TURING 图灵新知







下金蛋的数学问题



中国工信出版集团 人民邮电出版社 POSTS & TELECOM PRESS

韩雪涛◎著

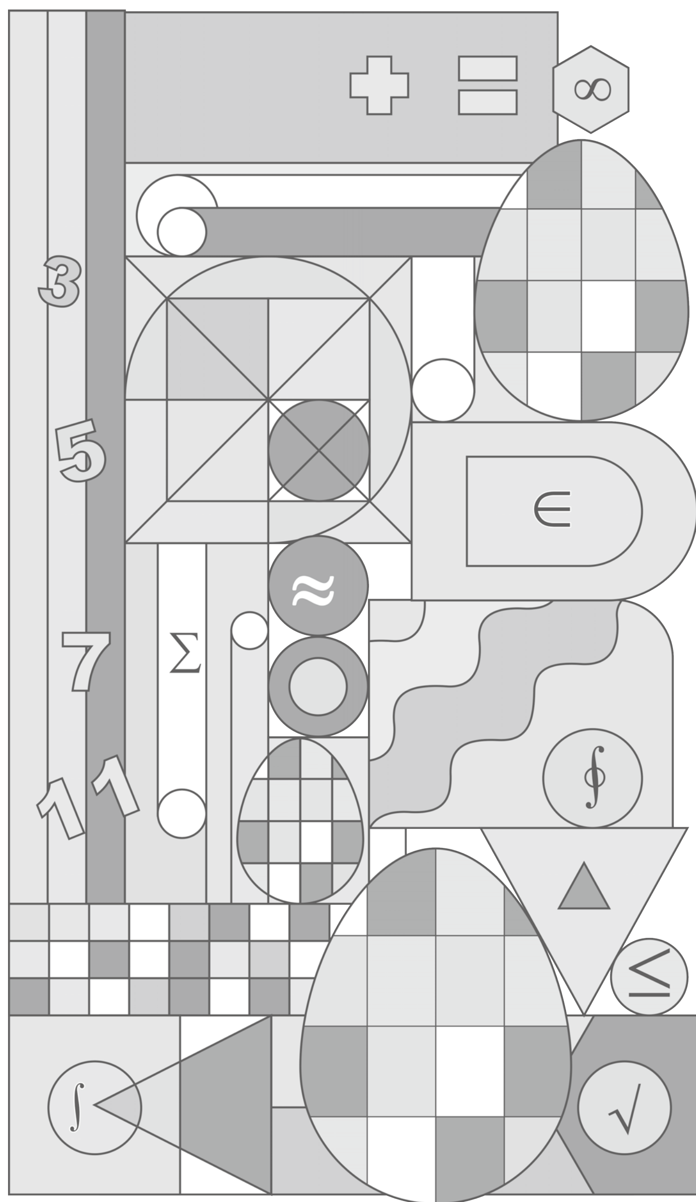


科普作家，另著有《数学悖论与三次数学危机》《从惊讶到思考——数学悖论奇景》等，参与编写《十万个为什么（第六版，数学卷）》《数学的足迹》（改变世界的科学丛书）等。

本书曾入选“2010年新闻出版总署向全国青少年推荐百种优秀图书”书目。《数学悖论与三次数学危机》在2016年入选《环球科学》最美科学阅读Top10书单。参与编写的《数学的足迹》在2016年10月荣获第四届中国科普作家协会优秀科普作品奖（图书奖）金奖。

数字版权声明

TURING 图灵新知



下金蛋的数学问题

韩雪涛◎著

人民邮电出版社
北京

图书在版编目（CIP）数据

下金蛋的数学问题 / 韩雪涛著. -- 北京 : 人民邮电出版社, 2020.7
（图灵新知）
ISBN 978-7-115-53836-9

Ⅰ. ①下… Ⅱ. ①韩… Ⅲ. ①数学—普及读物 Ⅳ. ①O1-49

中国版本图书馆 CIP 数据核字 (2020) 第 064348 号

内 容 提 要

本书介绍了从代数、几何、图论、数论中采撷出的 6 个经典数学问题。第一章介绍多项式方程根式解问题。第二章介绍几何三大问题，即用尺规三等分角、倍立方，以及化圆为方。第三章介绍欧几里得第五公设问题。第四章介绍四色问题。第五章介绍费马问题。第六章介绍素数问题。通过这几个问题的清晰介绍，读者可对这些问题的来龙去脉获得清楚认识。

另外，书中还穿插了数学家的逸事及相关的数学思想。通过这种介绍，读者可从更多侧面了解“数学家是什么样的人”，同时可对许多重要数学思想有更透彻的认识。

本书是一本数学科普读物，可供广大师生及所有数学爱好者阅读。

◆ 著　　　　韩雪涛
　责任编辑　王军花
　责任印制　周昇亮
◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
　邮编　100164　　电子邮件　315@ptpress.com.cn
　网址　https://www.ptpress.com.cn
　北京　　　　　印刷
◆ 开本：720×960　1/16
　印张：23.25
　字数：354 千字　　　　2020 年 7 月第 1 版
　印数：1 – 3 500 册　　　2020 年 7 月北京第 1 次印刷

定价：79.00 元

读者服务热线：(010)51095183 转 600　印装质量热线：(010)81055316

反盗版热线：(010)81055315

前　言

美国数学家哈尔莫斯说过：“问题是数学的心脏。”这一数学名言言简意赅地点明了数学问题对数学的重要性。

1900 年 8 月，在巴黎召开的第二次国际数学家大会上，德国数学家希尔伯特做了题为《数学问题》的著名演讲。这一演讲，成为世界数学史的重要里程碑，为 20 世纪的数学发展揭开了光辉的一页。在这次演讲中，希尔伯特提出了 23 个当时未解决的难题——后来以“希尔伯特问题”著称，而对这些难题的研究激励、推动了 20 世纪整个数学的发展。

1985 年，美籍华裔数学大师陈省身在南开大学数学研究所成立时指出，“一定要做好的数学”“有好的数学和不好的数学之分”“要从年轻时就懂得欣赏好的数学”。

从这两个事例中，我们可体会到好的数学问题对数学的重要性。那么，什么是好的数学问题呢？在陈省身看来，只有那些不断深入、有发展前途、可以影响许多学科的数学问题才是“好的数学”，如解方程。而另一些数学问题虽然可能也蛮有意思，但难以有进一步发展，它们就是“不好的数学”，如“拿破仑定理”。在希尔伯特看来，好的数学问题在于它有用且增进知识，数学史上的重要问题就在于其能创造新方法、建立新理论、开辟新领域。简单说，好的数学问题就是能为数学“下金蛋”的数学问题。

纵观数学发展史，这类重要的、有价值的数学问题可谓不胜枚举。本书所要介绍的正是从代数、几何、图论、数论中采撷出的 6 个经典数学问题。

- ❑ 在第一章中，我们介绍多项式方程根式解问题。这一问题涉及的是代数的中心问题：解方程。通过对这一问题的介绍，我们将看到代数学是如何随着这一问题的研究一步一步发展起来的。我们还将看到，正是问题最终的解决，将代数学引向了新的方向。

- ❑ 在第二章中，我们介绍几何三大问题，即用尺规三等分角、倍立方，以及化圆为方。这一类问题属于平面几何，而问题的解决要以解析几何作为工具。因此，我们在这一章中也会简单介绍一下解析几何。

- ❑ 在第三章中，我们介绍欧几里得第五公设问题。这一问题来自欧氏平面几何，但对它 2000 多年的探讨却促使了非欧几何的创立。我们还将看到，非欧几何的产生对数学的重要意义及其在相对论中的应用。

- ❑ 在第四章中，我们介绍四色问题。这一问题属于拓扑学，或更确切地说，属于图论。我们将看到，诞生于数学游戏的拓扑学与图论是如何随着四色问题的研究得到进一步发展的。最终四色问题的计算机证明，又引发了人们对数学证明等问题的深入探讨。

- ❑ 在第五章中，我们介绍费马问题。这一问题属于数论。我们的介绍亦将从数论的起源开始，并简单介绍在数论早期发展中做出重要贡献的几位数学家及其工作。最后，我们将以英国数学家怀尔斯的圆梦之旅作为这出精彩数学戏剧的尾声。我们还将从中看到，早期的数论伴随着这一问题的研究而得以扩展形成新的数学分支——代数数论。

- ❑ 在第六章中，我们介绍素数问题。这一同样属于数论的问题曾被列入希尔伯特 23 个问题中，也称为希尔伯特第 8 问题。这是一个涵盖面非常广的问题。

我们将主要介绍数学之圣杯：黎曼猜想。这一问题与本书前 5 章介绍的问题有一个重要差别：前者都是已经获解的问题，而只有黎曼猜想，这一被许多数学家认为是最重要的数学问题至今仍是尚未有人登顶的数学高峰。

通过这几个问题的清晰介绍，读者可对这些问题的来龙去脉获得清楚认识。而伴随着这一过程，读者的数学旅程亦将成为反复体验“从惊讶到思考”的快乐之旅。

首先，读者会为这几个著名难题拥有的“困难性和简单性的某种巧妙组合”特征而惊讶。简单性，是指问题本身清楚、易于理解。正是这一特点吸引、诱惑着无数人走上寻找问题解决钥匙的道路。同时，无数尝试者的失败也揭示出这些看似简单的问题实质上极度困难。这种问题表述“非常简单”和问题求解“极度困难”的鲜明对立，是这些问题的特别迷人之处，也是令人惊叹之处。

其次，当看到这些问题的最终解答时，读者可能有更大的惊讶感。一方面，有些问题的答案本身出人意料，如高于四次的一般多项式方程没有根式解。另一方面，更出人意料的是问题的解决途径。比如，为了证明看上去很简单的四色问题，我们不得不借助于电子计算机。再如，为了证明费马大定理，数学家们不得不兜上一个大圈子证明谷山 – 志村猜想，而为了证明谷山 – 志村猜想，怀尔斯不得不综合利用现代数学许多分支的成就。

最后，读者可能惊讶于这些数学问题孵出的金蛋之多。通过书中例证，我们会清楚认识到，数学中的重要问题，往往会成为新思想发展的酵母。为了解决问题，数学家们往往要提出新概念、新思想，创立新方法，得到新发现，而这些真知灼见，甚至可能打开一扇新分支的大门。比如，正是在研究多项式方程的根式解问题中，天才数学家伽罗瓦提出了群论。再如，在费马大定理的研究中，德国著名数学家库默尔提出了理想数思想，其新思想虽然没有彻底解决费马猜想，但极大推动了代数数论的发展。

从不断的惊讶中，读者不仅会对这些重要的数学问题有所认识，从中领略它们的

魅力，而且可深切体会“重大而关键的问题是活的血液，是推动数学发展的重要动力源泉”。更重要的是，读者在惊讶之余，还可做多角度、多层次的深入思考。

- 数学是一个不可分割的有机整体，它的生命力正在于各个部分之间的联系。我们从本书的介绍中会反复体会到：在一个领域中发展的成果，将成为解决另一个似乎无关领域中突出问题的关键因素。而这种稳固出现的联系证实了数学本质上的一致性。

- 一般来说，如果继续沿着前人的老路去解决著名数学问题，小改小革难以奏效，对于有几百年历史的老问题尤其如此。一个历史的经验是，要解决问题，必须找到新方法和新途径，特别是发现与其他领域的新关系。自然，想依靠初等方法解决这类著名难题的努力注定会是徒劳的。

- 我们还会体会到，抽象的与高深的数学是必要的。现代数学抽象化的趋势，不免引起人们对数学的误解，认为数学只是少数怪杰才能问津且远离现实的象牙塔。实际上，数学的抽象绝不是无源之水、无本之木。书中提到的群论的诞生即是很好的例证。

真正想从事数学研究，并有志解决世界数学难题的人，应该明白自己努力的方向：打下良好的基础，具备相当的数学知识与修养，掌握所研究领域和课题已有的成果、方法和最新文献，在此基础上，再做进一步的探讨。也就是说，任何世界难题的解决，都建立在对前人成功与失败的深入了解的基础之上。如果认为自己可以完全独立另造一片天空，以解决前人未能解决的问题，这只能是白费力气。

从惊讶到思考，我们将加深对“数学是什么”的理解，而且还可体会到数学之美，感受到数学的无穷魅力。

另外，书中还穿插了数学家的逸事及相关的数学思想。通过这种介绍，读者可从更多侧面了解“数学家是什么样的人”，同时可对许多重要数学思想有更透彻的认识。

本书是一本数学科普读物，可供广大师生及所有数学爱好者阅读。

本书在写作过程中参考了大量的数学书籍（书后附有主要的参考文献），谨向这些书的作者和译者表示真诚的谢意。另外，本书在写作过程中还参考了部分网上相关资料，书中部分图片也是通过搜索引擎在网上找到的，在此谨向这些网文与图片作者或所有者表示感谢。本书无法一一注明参考网文与图片的来源，还请见谅。

最后需要说明的是，书中不足或错误在所难免，我真诚地期望能得到读者朋友的指正。如果你有什么意见或建议，可以通过电子邮箱 zhhxt@163.com 与我联系。

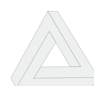
目　录

第一章
多项式方程根式解问题

第二章
几何三大问题

第三章
欧几里得第五公设问题

第四章 四色问题



第一章

多项式方程根式解问题

第一节
河谷文明与多项式方程

在本章中，我们将介绍多项式方程［即一元n次方程（以下简称n次方程），有时也称代数方程］求解（根）公式的探寻历程，这种公式要求通过对方程的系数进行有限次四则运算与开方运算，最终给出方程的解。由于方程的求解公式离不开根式，所以人们也把多项式方程的求解公式问题称为根式解问题。人类对此的最早尝试可追溯到遥远的古代文明。

历史学家往往把兴起于古埃及、美索不达米亚、中国和印度等地域的古代文明称为“河谷文明”，而早期数学就是在尼罗河、底格里斯河与幼发拉底河、黄河与长江、印度河与恒河等河谷地带首先发展起来的。从可考证的史料看，古埃及与美索不达米亚的数学在年代上更为久远，只是在公元前均告衰微，崛起稍晚的中国与印度的数学则延续到公元纪元之后并在中世纪臻于高潮。

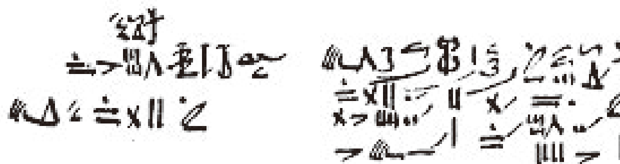
在本节中，我们将简单介绍古埃及与美索不达米亚这两个河谷文明在求解多项式方程方面取得的成就，中国和印度在这方面的贡献将放在第二节中做介绍。

古埃及人的成就

人们对古埃及数学知识水平的了解主要来自幸存至今的两部纸草书。一部叫《莱茵德纸草书》，因1858年为苏格兰收藏家莱茵德收藏而得名。它是在公元前1650年左右由一位名叫阿姆士（Ahmose）的僧侣抄录的，书的开头写道："准确的计算。阐明一切现有事物和模糊秘密的指南。"后世也以《阿姆士纸草书》作为书名。根据书的前言还知道，阿姆士并非书的原著者，他抄录的其实是一部已经流传了两个多世纪的更古老的著作。

另一部称为《莫斯科纸草书》，它在1893年为一位俄罗斯收藏者获得，1912年转为莫斯科博物馆所有，并因之得名。据研究，它出自约公元前1890年一位佚名作者之手。

这两部纸草书是古埃及最重要的传世数学文献。它们都是各种类型的数学问题集。《莱茵德纸草书》主体部分由84个问题组成，《莫斯科纸草书》则包括了25个问题。在这些问题中，有些可以归为今天所说的代数方程的范畴。我们剖析其中一个简单例子来看一下。



上图是《莱茵德纸草书》第24题的僧侣文原文，第一个字如下图所示：



读音类似“阿哈”（aha）或“呵”（hau），意思是“一堆”，相当于方程中的未知数。整个题目可意译为：已知“堆”与七分之一“堆”相加为19，求“堆”的值。

古埃及人对此问题的解法比较有趣。他们先把7作为未知数的实验值，代入得数8，但应得结果是19，这两个结果之比是$\frac{19}{8}$，于是将7乘以$\frac{19}{8}$即得出正确的答案。

《莱茵德纸草书》第26个问题也使用了这种技巧。这个问题可意译为：求一未知数，它与自身的$\frac{1}{4}$相加的结果为15。纸草书中的解法如下：假设答数为4，那么4加4的$\frac{1}{4}$为5……找一个乘以5能得到15的数，答案是3，再用4乘以3，答案是12。

这种解法被称为“假位法”，实质上是一种算术方法：先假设一个特殊的数作为“堆”值，将其代入等号左边去运算，然后比较得数与应得结果，再通过比例方法确定未知数的真值。这种假位法是《莱茵德纸草书》中普遍使用的方法，因其过程采用了一次假设，故这种解法也叫“单假设法”。

对此，著名数学史家史密斯曾评述说：“世界曾经为形如$ax+b=0$的方程所困惑过，这似乎是不可思议的，但是古代数学家为解这种方程，确实曾求助于一种比较烦琐的方法，这种方法后来在欧洲称为‘假位法’。”

不过，古埃及纸草书中也记载了用我们现代的方法来求解一次方程。例如，《莫斯科纸草书》中有一个例子是：求一个数，它的$1\frac{1}{2}$倍加上4等于10。用现代的记号表示即$1\frac{1}{2}x+4=10$，其解法是：首先10减去4，然后将6乘以$\frac{2}{3}$，得解为4。

在纸草书中，古埃及人还求解了形如$x+ax=b$或$x+ax+bx=c$的一次方程。除了含有一个未知数的一次方程问题外，古埃及人还研究并处理了简单的二次方程。如在别的纸草书中有一个问题：把一个面积为100的正方形分为两个小正方形，使其中一个的边长是另一个的$\frac{3}{4}$。这个问题用现代的记号表示为：

$$\begin{cases} x^2+y^2=100 \\ x:y=1:\frac{3}{4} \end{cases}$$

我们看到，从这一问题可引出一个二次方程。有意思的是，古埃及人对此的处理也是使用假位法。其解题思路是：先假设 $x=1$，则 $y=\frac{3}{4}$，$x^2+y^2=\left(\frac{5}{4}\right)^2$，右端是边长为 $\frac{5}{4}$ 的正方形面积，而题目中大正方形的边长是 10，两者的比是 $10:\frac{5}{4}=8$，因此将原假设的边长扩大 8 倍，得 $x=8$，而 $y=8\times\frac{3}{4}=6$，此即为原问题的正确答案。

可以看到，古埃及人解决了相当于 $ax^2=b$ 类型的最简单的二次方程。

古巴比伦人的成就

美索不达米亚文明（其典型代表是古巴比伦文明）大约在公元前 3500 年出现于底格里斯河和幼发拉底河流域。人们对其的了解来自保存至今的约 50 万块泥板。

古巴比伦人曾用尖芦管在湿泥板上刻写楔形文字，然后将泥板晒干或烘干。后来，人们挖掘出大量这种不易毁的泥板。这些被保存下来的泥板中有 300 多块属于数学文献。它们主要分属两个相隔遥远的时期。一大批是公元前 2000 年头几个世纪的遗物，还有许多则来自公元前 1000 年后半期。通过研究这些泥板文书，人们发现古巴比伦人在多项式方程求解方面表现出高度的技巧，取得了十分可观的成就。

对于一次方程，古巴比伦人或许觉得简单，没有过多探讨。在留下的资料中很少有这类方程，并且没有提供求解算法。如一块泥板上有这样的问题：一块石头，加上它的 $\frac{1}{7}$，得到一个重量，在这个重量之上再加它的 $\frac{1}{11}$，得到 1 马那（mana，重量单位），问这块石头有多重？这个问题相当于方程 $\left(x+\frac{x}{7}\right)+\frac{1}{11}\left(x+\frac{x}{7}\right)=1$，可解得 $x=\frac{77}{96}$

马那。答案是正确的，但没有解题的过程。

我们已看到古埃及人主要讨论的是一次方程，对于二次方程则仅涉及最简单的情形 $ax^2=b$。与之不同，古巴比伦人的突出成就恰恰表现在求解二次方程方面，这也是他们的一个主要研究领域。事实上，许多古巴比伦泥板上列有大量的二次方程。例如一块泥板文书中有这样的问题：已知 igibum（依几布姆）比 igum（依古姆）大 7，问 igibum 和 igum 各为多少？

这里的 igibum 和 igum 是古巴比伦数学文献中表示互为倒数（即乘积为 1）的两个数的专有术语。值得注意的是，古巴比伦人在这里相当于用记号表示了未知量。另外需要说明的是，古巴比伦人使用 60 进制，igibum 与 igum 乘积为 1，1 指一个大单位，相当于 60 个小单位，而问题中的 7 则是小单位。于是若以 x 表示 igibum，y 表示 igum，则上述问题相当于求解方程组：

$$\begin{cases} xy=60 \\ x-y=7 \end{cases}$$

如我们所熟知的，这可转化为求解一个二次方程 $x^2-7x-60=0$，古巴比伦人给出算法：$x=\sqrt{\left(\frac{7}{2}\right)^2+60}+\frac{7}{2}=12$。这相当于对二次方程 $x^2-px-q=0$ 使用了求根公式 $x=\sqrt{\left(\frac{p}{2}\right)^2+q}+\frac{p}{2}$，这与我们更熟悉的 $x=\frac{p+\sqrt{p^2+4q}}{2}$ 显然是等价的。

古巴比伦人也给出了另一类似问题——求一个数，使它和它的倒数之和等于一个给定的数——的正确解法。

进而，古巴比伦人还研究并解决了更具一般性的一种标准形式：已知两数的和与积，求这两个数。对要求的两个未知量，他们使用的单词是长（us）、宽（sag），对两者的积使用的单词是面积（asa）。所要解决的问题与问题的陈述都使人想到，古巴比伦人想处理的是矩形的面积、周长与边长之间的关系。事实上，他们研究的这种标准形式

相当于：已知矩形的周长与面积，确定矩形的长与宽。这一问题的求解在古代既重要又迫切，因为当时人们普遍有一种错误认识：矩形的面积完全取决于它的周长。许多懂得多的人就利用这一点欺骗他人，在占便宜的同时还赢得了慷慨的名声——他们自己拿到周长较小面积较大的土地，给别人周长较大面积较小的土地。

不过，随着时间的推移，这一问题逐渐失去几何意义，“长”“宽”这些几何术语也慢慢成为表示未知量的标准符号，并演变成抽象的代数符号，相当于我们熟悉的x和y。因此，古巴比伦人研究的这种标准形式用现代的代数语言来叙述，那就是：已知$x+y=p$, $x\cdot y=q$，求x, y。

古巴比伦人用下述五个步骤求这两个数：

(1) 取p的一半；

(2) 将此数平方；

(3) 从中再减去q；

(4) 对所得结果开平方；

(5) 再加p的一半得出所求两数中的一数，从p中减去这个数得出另一个数。

容易明白，他们所解的方程组可转化为二次方程：$x^2-px+q=0$。而他们的五步解法，用现代代数语言来写就是：$x=\sqrt{\left(\frac{p}{2}\right)^2-q}+\frac{p}{2}$，$y=p-x$。这与我们更熟悉的形式$x=\frac{p+\sqrt{p^2-4q}}{2}$和$y=\frac{p-\sqrt{p^2-4q}}{2}$自然是等价的。

除求解由方程组获得的二次方程外，古巴比伦人也求解单个的二次方程。如有这样一个问题：一正方形的面积与它的一条边的$\frac{4}{3}$之和为$\frac{11}{12}$，求其边长。

用现代的术语来说，就是要解方程$x^2+\frac{4}{3}x=\frac{11}{12}$。古巴比伦人求解的步骤与方法是：取$\frac{4}{3}$的一半，得$\frac{2}{3}$，取$\frac{2}{3}$的平方，得$\frac{4}{9}$，然后将这个结果与$\frac{11}{12}$相加，得$1\frac{13}{36}$，这

个值是$\frac{7}{6}$的平方。从$\frac{7}{6}$减去$\frac{2}{3}$得$\frac{1}{2}$，这就是所需要的边长。这一算法可以翻译成现代求解方程$x^2+bx=c$的公式，即$x=\sqrt{\left(\frac{b}{2}\right)^2+c}-\frac{b}{2}$。可以看出，这是二次方程求根公式的一种解形式。

还有一个问题如下：如果某正方形的面积减去某边长得870，问边长是多少？用现代的代数语言表示，即解方程$x^2-x=870$。泥板给出的解法是：取1的一半，得$\frac{1}{2}$；以$\frac{1}{2}$乘以$\frac{1}{2}$，得$\frac{1}{4}$；把$\frac{1}{4}$加在870上，得$\frac{3481}{4}$，它是$\frac{59}{2}$的平方，$\frac{59}{2}$再加上$\frac{1}{2}$，结果是30。这一算法可以很容易地翻译成现代求解方程$x^2=bx+c$的公式，即$x=\sqrt{\left(\frac{b}{2}\right)^2+c}+\frac{b}{2}$。可以看出，这也是二次方程求根公式的一种解形式。

此外，古巴比伦人还研究并给出了形如$x^2+c=bx$的方程的正确解法。

可以看到，古巴比伦人把不缺项二次方程分成了不同的形式并分别求解，现代读者可能会对此深感困惑。古巴比伦人为什么不像我们现在所做的这样把二次方程统一处理，而非要自找麻烦呢？原因在于，古代没有负数的概念。因此，未知数的系数、常数及方程的解都必须是正数。这样，要研究的不缺项二次方程就被分成三种基本类型：$x^2+bx=c$、$x^2=bx+c$、$x^2+c=bx$（其中b、c均为正数），而且要对不同的类型分别求解。通过后面的介绍，我们会明白这种分类处理方程的方式在数学发展上延续了相当长的时期。

所有这三类方程在古巴比伦泥板文书中都可以找到，并都给出了正确的解算程序。而古巴比伦人也清楚如何把给定的二次方程转化成上面三类方程之一。可以说，古巴比伦人已经圆满地解决了二次方程的求解问题，这在古代特别是几千年以前是相当非凡的成就。不过，需要指明的是，古巴比伦人并没有真正地使用类似如今多项式方程的概念，其解法也都是通过文字来叙述求解步骤。事实上，他们所谓的算法是由一个个简练的短语或句子组成的，没有“等号”或其他简洁符号。自然，因为没有符号表示法，所以也不可能给出我们现在所使用的一般二次方程的求根公式，他们所能做的

只是给出具体例题，并通过这些具体例题的求解来说明一般解法。也正是由于没有符号表示法，他们在解题时才会遇到那么多困难。这样的困难是代数语言系统发展起来以前人们不得不面对的。

此外，古巴比伦人虽给出了求解的正确步骤，却没有说明每一步做法的理由。后人对此做出一些推测，比如“已知 $x+y=p$，$x\cdot y=q$，求 x,y”，古巴比伦人的思路可能如下：若取 $x=\frac{p}{2}$ 和 $y=\frac{p}{2}$，则满足 $x+y=p$，但不一定满足 $x\cdot y=q$。因此，这解应有误差，假设误差是 z，则可设真正的解 $x=\frac{p}{2}+z$ 和 $y=\frac{p}{2}-z$。于是 $x\cdot y=\left(\frac{p}{2}+z\right)\left(\frac{p}{2}-z\right)=\frac{p^2}{4}-z^2=q$。由此可以得到 $z=\sqrt{\left(\frac{p}{2}\right)^2-q}$。进而，可得出 x,y。

除了有效地解决一般的不缺项二次方程外，古巴比伦人还处理了古埃及人未探讨的三次方程。像最简单的 $x^3=a$，主要通过查立方表或立方根表来求解。事实上，为了使计算简洁，古巴比伦人编制了大量的数学用表。在现有的 300 多块数学泥板文书中，就有 200 多块是数学用表，包括乘法表、倒数表、平方表、立方表、平方根表、立方根表，甚至还有指数表、对数表。对形如 $x^3+x^2=a$ 的混合三次方程，他们也是借助现成的表来求解，因为古巴比伦人编有专门的 n^3+n^2（其中 n 为整数）的数值表。

对一个特殊三次方程 $144x^3+12x^2=21$ 的解决，显示了他们在解方程方面的高超水平。古巴比伦人先运用了代换的方法：用 12 乘以方程两端，并设 $y=12x$，从而把方程转化为 $y^3+y^2=252$，查表得 $y=6$，因此 $x=\frac{1}{2}$。在没有现代符号的情况下，能够认识到方程 $(ax)^3+(ax)^2=b$ 与 $y^3+y^2=b$ 本质上属于同一类型，这种初等的代数变换思想在当时是了不起的成就。

古巴比伦人在研究中还涉及并解决了简单的四次方程。比如泥板中有这样一个问题：“我把长乘宽得面积 10。我把长自乘得面积，我把长大于宽的量自乘，再把这个结果乘以 9。这个面积等于长自乘所得的面积。问长和宽是多少？”

如我们上面已经介绍到的，这里的文字长、宽，只不过是两个未知量的方便说法，

相当于我们的 x、y。于是这个问题用现在的符号可表示为：

$$\begin{cases} xy = 10 \\ 9 \times (x-y)^2 = x^2 \end{cases}$$

在解这一方程组时，会得到 x 的一个四次方程，但其中缺少 x 和 x^3 项，因而这个方程可作为 x^2 的二次方程来解。古巴比伦人给出了这一问题的正确答案。

通过上述的介绍，我们看到古巴比伦人可以圆满解决一次、二次方程，可以求解部分特殊的三次、四次方程，古巴比伦人在多项式方程研究方面取得的成就在古代数学中确实是非常突出的。



第二节
两位代数学之父

对于比较简单的多项式方程的求解方法，许多古代民族都在探索中得到了正确的结果。我们在上一节中介绍了古埃及人，特别是古巴比伦人在这方面取得的引人注目的成就。在本节中，我们将简要介绍一下古代中国与古印度在多项式方程求解中取得的成就。

此外，多项式方程的求解与代数学的发展有着极为紧密的联系。事实上，直到 19 世纪，求解 n 次方程一直是代数学研究的中心课题。因此，我们将在本节中介绍两位“代数学之父”，并着重介绍他们在多项式方程求解方面的贡献。

古希腊的丢番图与《算术》

丢番图（约 200—284），古希腊著名数学家，曾经居住在埃及的亚历山大城。一本《希腊诗文选》（约公元 500 年）收录了他奇特的墓志铭——关于他的生平事迹，后人主要是通过这一别开生面的记载了解的。

坟中安葬着丢番图，
多么令人惊讶，
这里忠实地记录了他所经历的道路。
上帝给予的童年占六分之一，
又过十二分之一，两颊长胡，
再过七分之一，点燃起结婚的蜡烛。
五年之后天赐贵子，
可怜迟到的宁馨儿，
享年仅及其父之半，便进入冰冷的墓。
悲伤只有用算术的研究去弥补，
又过四年，他也走完了人生的旅途。

由这一诗文，可得到方程 $\frac{1}{6}x+\frac{1}{12}x+\frac{1}{7}x+5+\frac{1}{2}x+4=x$，解这个一次方程，得 $x=84$，由此可知丢番图享年 84 岁。

丢番图最重要的作品是《算术》。作为一部划时代的著作，它在历史上影响之大可以和欧几里得《几何原本》相比。《算术》最初有 13 卷，但没有完整保存下来。很长时间里，人们只见到它的 6 卷希腊文本。直到 1973 年，人们才奇迹般地发现了 4 卷先前不为人知的阿拉伯文译本。丢番图在数学上的地位基本上就是靠这并不完整的《算术》赢得的。

丢番图的著作实质上是一长串问题集，其中希腊文本有 189 个，阿拉伯文本有 101 个，共收录了 290 个题目。此外，还有十几个引理和推论，合起来共 300 多个问题。这些问题大都相当于现在“数论”（后面章节会做介绍）的内容。但此著作也涉及代数问题与代数思想，我们现在简单介绍一下。

对于一次方程，丢番图可能觉得太简单，没有必要单独论述，因此直接给出问题

的答案。但他熟悉现在我们所使用的“移项”“合并同类项”等操作，由此可以推断，他解一次方程的思路与我们现在的解题思路是一致的。

其著作中出现了若干二次方程或可归结为二次方程的问题，这些例子足以说明丢番图熟练掌握了各种类型的二次方程的求解方法。但书中现存章节并没有给出二次方程的系统解法，所以人们不知道他是用什么方法解的。

值得一提的是，丢番图已经懂得负数运算的符号法则，他对负数的认识是超越时代的。不过，丢番图对负数解是排斥的。答案是 0 时，他也舍弃。对无理数解他也认为不合理。简单地说，丢番图的解答只限于正有理数解。有意思的是，对他来说，找到问题的一个答案就够了。当方程有两个正根时，他始终只取较大的一个。

可以说，在书的编排（问题集）与解题技巧等方面，丢番图与古巴比伦人采用了颇为相似的方式。因此，他的工作有时被说成是“盛开的巴比伦代数的花朵”。但在符号体系方面，丢番图做出了更重要的突破。

在第一节中，我们看到古代人在处理代数问题时，所有内容都用文字表述，就像写散文一样。在这一被称为“文辞代数”的阶段，除了偶尔出现的表示未知量的符号外，几乎不使用任何代数符号。与前人不同，丢番图开始有意识地使用符号体系。

我们来看一下他所创设的一些符号。比如说，他用 $\overset{o}{M}$ 表示数的单位，取自希腊字“单位”的字头；未知数用特殊的符号 ς 表示；“未知数的平方”记作 Δ^{Υ}，取自希腊字“幂”（ΔΥΝΑΜΙΣ）的前两个字母。同样，“未知数的立方”记作 K^{Υ}，取自希腊字“立方”（ΚΥΒΟΣ）的前两个字母。此外，他用 $\Delta^{\Upsilon}\Delta$ 表示平方的平方，ΔK^{Υ} 表示平方的立方，$K^{\Upsilon}K$ 表示立方的立方。对这些乘幂的倒数，丢番图也给出了专名和符号。在他的符号体系中没有加号、乘号、除号，但特别给出了减法的符号 ⋔。

在写方程时，未知数的系数紧接在未知数后面，加通过并列来表达，所有的负项都放在减号的后面，若有常数项，就在希腊字“单位”的字头 $\overset{o}{M}$ 后加上系数。举一个

简单的例子，式子 x^3+2x-3 用丢番图的符号体系可表示为：$\mathrm{K}^{\mathrm{T}}\alpha\varsigma\beta\pitchfork\mathring{\mathrm{M}}\gamma$，其中 α、β、γ分别表示数字 1、2、3。

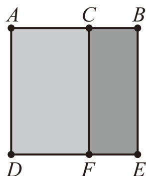
可以看到，丢番图的符号多半采自相应文字的字头。这种以速记形式的缩写记号来表示经常出现的数量和运算的方式被称为缩写代数。当然，丢番图的缩写代数并不具有现代代数方程的简洁形式，不是特别容易读懂。然而，与前人相比，他在发展专门的符号体系方面毕竟向前迈进了一大步，将代数记号的发展推进到一个新的阶段：缩写代数。

在历史上，丢番图第一次系统地提出代数符号，这一努力为代数带来了重大变革，推动了代数学的发展。正如数学史家克莱因所说：“代数的进步是因为引用了较好的符号体系。对于它本身和分析的发展来说，这比 16 世纪的技术进步远为重要。事实上，采取了这一方式，才使代数有可能成为一门科学。在 16 世纪前，自觉运用一套符号以使代数的思路和书写更加紧凑、更加有效的人只有丢番图。”

丢番图对代数学发展做出的另一划时代贡献是，他将代数从几何形式中脱离出来，使它成为数学中的一个分支。

在丢番图之前，古希腊数学的兴趣集中在几何学上。古希腊人认为只有经过几何论证的命题才是可靠的。于是，所有的数都以几何图形的量来表示，而且所有的代数问题都被改写成几何问题来处理。充分体现古希腊代数这一特点的有《几何原本》第二卷中的命题一，我们以此为例来看一下。

命题一：设有两线段，其中之一被截成若干部分，则由此两线段所构成的矩形的面积等于各个部分与未截线段所构成的矩形的面积之和。



若设 $AD = x, DF = y, FE = z$ ，则这一几何命题用符号可以表示成 $x(y + z) = xy + xz$ 。因此，这一命题表述的意思其实相当于我们所熟知的乘法分配律。

古希腊这种用几何语言来表述的、可视化的代数学被称为几何代数学。很长时间里，古希腊人为了逻辑的严密性，给代数披上几何的外衣，把一切代数问题都纳入僵硬的几何模式之中。比如说，古希腊人曾借助这种几何方法间接求解了一些类型的二次方程。直到丢番图时代，代数才解放出来，摆脱了几何的羁绊，成为“自由人”，不再是几何的“仆从”。

最早使用代数符号体系，致力于方程的研究，把代数学从几何学中独立出来……由于对代数学发展所做出的这些划时代贡献，丢番图被一些数学史家誉为“代数学之父”。他的《算术》作为代数最早的巨著，代表了古希腊代数思想的最高成就。

中国古代数学中的代数方程

在四大河谷文明中，同样有着悠久历史的中国数学持续繁荣时间最为长久。从公元元年左右至 14 世纪，中国古代数学先后经历了三次发展高潮，即两汉时期、魏晋南北朝时期和宋元时期，并在宋元时期达到顶峰。在这一漫长的过程中，中国数学在多

项式方程求解方面取得了众多进展。但与我们本章要介绍的寻求求根公式不同，中国古代数学因为倾向于实用，所以热衷于方程的数值求解，并在这方面取得了杰出的成就。我们这里只简略提一下。

成书于 1 世纪的经典著作《九章算术》的“少广章”中最早记载了开平方术和开立方术。用现代记号表述，开平方术相当于解方程 $x^2 = A$，开立方术相当于解方程 $x^3 = A$。在这种开方过程中会遇到开不尽的情形，即遇到无理数。对此，中国古代数学家已能计算出这种不尽根数的近似值。魏晋时期的著名数学家刘徽在为《九章算术》作注时，更是明确提出“求微数法”，相当于用十进制小数任意逼近不尽根数。在《九章算术》的开平方术中，实际上还包含了二次方程 $x^2 + bx = c$ 的数值求解程序，它被称为“开带从平方”。到唐代时，数学家王孝通在《缉古算经》中研究了“开带从立方”，书中他给出 28 个形如 $x^3 + px^2 + qx = c$ 的正系数方程，并得出其正有理根。这表明，王孝通已掌握了求三次方程正根的数值解法。这些方法在宋元时代发展为一般高次方程的数值求解。先是在 11 世纪，宋代数学家贾宪创造出一种新的开方法——增乘开方法，通过随乘随加导出减根方程，逐步求出高次方程的正根。其后，12 世纪北宋数学家刘益第一次突破方程系数为正的限制。在前人工作的基础上，南宋数学家秦九韶（约 1202—1261）总其大成，提出一套完整的逐步求出高次方程正根的程序“正负开方术”，用这种方法可以简便有效地求出形如 $a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ 的高次方程的正根。这种求高次方程正根的一般方法现被称为秦九韶法，它与英国数学家霍纳 1819 年创立的霍纳法基本上一致。秦九韶法的提出，表明中国古代数学已经能解决任意次方程的数值求正根问题。

在求根公式方面，中国古代数学也获得了一些结果。公元 3 世纪三国时期的赵爽在为《周髀算经》作注时撰写了著名的《勾股圆方图注》，其中的一段文字相当于对形如 $x^2 + c = bx$（系数为正）的方程给出了求根公式。12 世纪数学家刘益在《议古根源》一书中，给出了另一类二次方程 $x^2 + bx = c$（系数为正）的求根公式。

此外，中国古代数学发展到宋元时期时，还出现了一个深刻的转向，即由“文辞

代数”阶段向代数符号化阶段转变。早期的中国古代数学处于“文辞代数”阶段，这在《九章算术》中有非常典型的体现。在这本中国古代数学经典著作中，所有的内容——问题、解以及求解方法——都只用文字和数，而不用数学符号表示，其中没有“等号”，没有表示未知量的x，也没有现在代数学使用的任何符号工具。但在宋元时期的数学著作中，开始做出了用特定的汉字作为未知数符号并进而建立方程的系统努力，这就是以李冶为代表的“天元术”和以朱世杰为代表的“四元术”。所谓“天元术”，就是设未知数布列方程的一般方法。首先是“立天元一为某某”，“天元一”就表示未知数，相当于现在的x，这相当于“x设为某某”，然后根据问题给出的条件列出两个相等的多项式，再通过所谓的“同数相消”，把这两个多项式相减，得到一个一端为零的一元方程式。用天元术列方程的方法，与现代代数中的列方程法非常类似。该方法后被元朝数学家朱世杰推广到最多有 4 个未知数的情形，即“四元术”。

李冶总结的天元术和朱世杰创立的四元术可看作是一种半符号代数，当然，这只是一种很不完备的符号代数。我们会在后面的介绍中看到，直到 16 世纪，在法国数学家韦达等人较系统地引入字母来表示数及常用的数学符号后，完善的符号代数才逐渐建立起来。

古印度数学中的代数方程

发源于印度河流域的印度文明，其历史可追溯到约公元前 3000 年，其数学发展的鼎盛繁荣时期是 5~12 世纪。其间，印度出现了许多有代表性的数学家，如婆罗摩笈

多、婆什迦罗等。我们只简略介绍这一时期他们在多项式方程方面取得的杰出成就。

婆罗摩笈多（约 598—660）在其主要著作《婆罗摩笈多修正体系》中，给出了一次方程与二次方程的解法。对于一次方程，他给出的法则是：绝对项之差变号后，除以未知数（系数）之差就是方程的未知数（值）。这一法则相当于，一次方程 $ax+b=cx+d$ 的解为 $x=\frac{-(b-d)}{a-c}$。对于二次方程，他给出“消去中项法则”：从平方项及简单项的另一端取绝对项。绝对项乘以平方项（系数）的 4 倍，加上中项（系数）的平方，（其和）的平方根，减去中项（的系数）把结果除以平方项（系数）的 2 倍，得中项（值）。这一法则相当于，二次方程 $ax^2+bx=c$ 的解为 $x=\frac{\sqrt{b^2+4ac}-b}{2a}$。

前面介绍中我们提到，古巴比伦人以及古希腊的丢番图等只考虑系数为正数的方程，这使得他们不得不把方程分成几种特殊情况来讨论。与之形成鲜明对比的是，婆罗摩笈多对系数有着十分深刻的理解，方程中的系数可为正数、负数或零。因此，婆罗摩笈多能像我们一样考虑方程的统一形式，得出一种更普遍、更现代、更有力的求解多项式方程的方法。此外，方程的解可以是正数，也可以是负数，可以是有理数，也可以是无理数。这都是相当先进的思想。

婆罗摩笈多的著作还有一个显著的特点，即代数符号的引入与应用。与丢番图的代数学一样，他的代数学也是简写代数。例如，他在一个数上面加一点来表示负数。当方程含有一个或多个未知量时，他称每个未知量为一种不同的颜色，并用颜色单词的缩写形式来表示未知数。他所使用的方法与代数学用字母 x、y、z 表示变量完全类似。采用缩写文字和符号来表示未知数和运算是印度人对代数学的一大贡献。

古印度最伟大的数学家、天文学家婆什迦罗（1114—1185），继承并发展了婆罗摩笈多的工作。他对一次和二次方程做了更详尽的讨论，在其著作中还有一个三次方程和一个双二次方程的例子。特别是在二次方程上，他所做的工作与现在中学生所学的完全一样。正如我们所介绍的，各民族数学家为了走到这一步，花费几千年的时间去

拓展数的概念、解的概念，以及解这些类型方程的计算技巧。

简单地说，古印度人在其数学的繁荣时期把多项式方程的求根问题推到了相当高的水平，并对代数学的发展做出了重大贡献。公元 8 世纪时，婆罗摩笈多的著作被带到巴格达，并被译成阿拉伯文，对当时阿拉伯的天文学和数学产生了一定影响。

古阿拉伯的花拉子密与《代数学》

阿拉伯文明是后起之秀。公元 7 世纪初，阿拉伯民族开始崛起。公元 8 世纪，阿拉伯数学开始进入兴盛繁荣期。当时的阿拉伯学者们广泛吸收希腊、印度与中国的数学成果，使阿拉伯数学在融合东西方数学的基础上发展出既含有西方因素也含有东方因素的折中数学。

当时阿拉伯最重要的数学活动中心在巴格达，阿拔斯王朝于 830 年在那里设立了著名的“智慧宫”，聚集了大批学者，其中最卓越的代表是花拉子密（约 783—850）。813 年，花拉子密被聘到巴格达工作，智慧宫建立后他成为主要领导人之一。在数学方面，他完成了两部传世之作：《代数学》和《印度的计算术》。其中，《代数学》完成于 820 年左右，是花拉子密影响最大的著作，被誉为代数发展史上的里程碑。



花拉子密

《代数学》全书共三部分。在第一部分，他讲述了现代意义下的初等代数，系统讨论了一、二次方程的解法。

花拉子密先指出这些方程由三种量组成：根（指未知数，把未知数叫根始于花拉子密，把解方程求未知量叫作求根也来源于此）、平方（指未知数即根的平方）、数（指常数）。然后，花拉子密把所讨论的方程分为六种类型。

前三种类型分别是：“平方”等于“根”，“平方”等于“数”，“根”等于“数”。用现代符号可分别表示为：$ax^2=bx$，$ax^2=c$，$bx=c$。容易看到，这三种类型表示的是缺项的二次方程与一次方程。

后三种类型分别是：“平方”和“根”等于“数”，“平方”和“数”等于“根”，“根”和“数”等于“平方”。用现代符号可分别表示为：$ax^2+bx=c$，$ax^2+c=bx$，$bx+c=ax^2$。可以看到，这三种类型表示的是不缺项的二次方程。

为什么不像我们现代所做的，把二次方程统一表示为$ax^2+bx+c=0$（其中$a\neq0$）呢？为什么要区分出这么多种类型？原因仍然在于，当时人们还不接受负数，所以方程的系数必须全取正数。因此，上述六种类型中出现的系数a、b、c全都要求是正数。

这意味着，我们可以统一解决的二次方程，花拉子密却不得不分成各种不同类型分别进行处理。由此，即可体现出数学中引入负数的必要性。

花拉子密对每一种类型通过举例题的方式给出了求解的方法。令现代人非常陌生的是，花拉子密完全用文字来表述问题及求解，没有用任何字母和缩写符号。比如在第四章“平方和根等于数”中，花拉子密写道：“下面是‘平方和根等于数’的一个例子：1 个平方和 10 个根等于 39。因而，在此方程中的问题是：当平方是多少时，它与其 10 个根之和是 39？解这类方程的方法是，先求出问题中提到的根数的一半，这个问题中根数是 10，因而得到 5，将其自乘，得 25，把它同 39 相加，得 64，开平方得 8，从中减去根数之半即 5，余 3。数 3 就表示所求平方的一个根。当然，平方本身是 9。因而所要求的平方是 9。”

很难理解，是不是？其实，换用现代符号，花拉子密举的例子可表示为：$x^2+10x=39$。其求解的步骤相当于：

$$10\times\frac{1}{2}=5, 5^2=25, 25+39=64, \sqrt{64}=8, 8-5=3, x=3, x^2=9$$

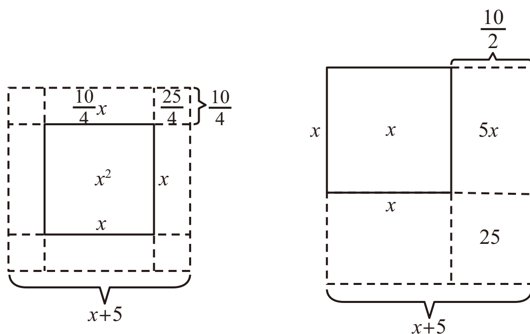
用简洁的式子表示为：$x=\sqrt{\left(\frac{10}{2}\right)^2+39}-\frac{10}{2}$。容易看到，这一求解过程与我们用求根公式解是一样的。但由于没有引入字母系数，对这种可以应用于$x^2+bx=c$一般形式的解法，花拉子密只是给出了一个特例做说明。

如果出现 $ax^2+bx=c$，如何处理呢？花拉子密的办法是，方程两边同除以 a，转化成$x^2+\frac{b}{a}x=\frac{c}{a}$，这是已经解决的形式，于是原方程的解用现代符号可表示为：$x=\sqrt{\left(\frac{b}{2a}\right)^2+\frac{c}{a}}-\frac{b}{2a}$，与我们所熟悉的结论是相同的。

花拉子密在书中以类似的方式，继续讨论了其他几种类型方程的解法。在充分讨论了六种典型方程的解法之后，深受希腊几何学影响的花拉子密觉得“有必要用几何学方式来证明曾用数字解释过的问题的正确性”，于是他又用几何方法给出它们的证

明。以上述所讨论的方程$x^2+10x=39$为例，花拉子密给出两种几何证明。

第一种证明是，在边长为x的正方形的四个边上向外作边长为x和$\frac{10}{4}$的矩形，再在四个角上补上边长为$\frac{10}{4}$的四个小正方形，从而把这个图补为边长是$x+5$的大正方形，如下面的左图所示。第二种证明是，在边长为x的正方形的两个相邻边上作边长为x和$\frac{10}{2}$的矩形，再补上一边长为$\frac{10}{2}$的小正方形，最终把图形补充为完整的大正方形，如下面的右图所示。



两种情况下，都可由已知得到大正方形的面积是$x^2+10x+25=39+25=64$，从而$x+5=8$，$x=3$。

在几何证明之后，花拉子密又明确提出了解方程需要的两种基本变换——“还原”（相当于现在的“移项”）与“对消”（相当于现在的“合并同类项”）。他指出，经过这两种变换，一般形式的一次和二次方程就都能转换成已经讨论过的六种标准方程。

在对方程根的了解方面，花拉子密最早认识到二次方程有两个根。不过，他只给出了正根，负根与零根则舍弃。他也认识到，当根的数目之半自乘的结果小于自由项时（相当于我们所说的判别式为负），开平方是不可能的，此时方程无根。

通过以上介绍，我们可看到这一著作中存在的缺陷：其一，不承认负数，方程的系数与方程的根都只取正数；其二，全部内容都是用语言文字来叙述的，没有使用代数符号，这与丢番图和古印度数学家相比都倒退了一步。

虽然如此，花拉子密毕竟完整解决了一次与二次方程的求解问题，并且其阐述条理清晰、详尽系统，使读者容易掌握解题方法。从那以后，解方程的概念逐步明朗起来，独立的代数学科亦由此确立下来，而花拉子密的工作则为这门新学科提供了方向。从这时起，解方程作为代数基本特征被长期保持下来。简言之，花拉子密的《代数学》奠定了以方程论为中心的古典代数学的基石，为“解方程的科学”的代数学开辟了道路。事实上，这门新学科的名称——代数学——本身就来源于此书。

《代数学》的阿拉伯文书名是 ilm al-jabr wa'l muqabalah，直译应为“还原与对消的科学”。al-jabr 意为“还原”，这里指把负项移到方程另一端“还原”为正项，相当于我们现在所说的“移项”；muqabalah 意即“对消”或“化简”，指方程两端可以消去相同的项，相当于我们现在所说的合并同类项。这个长长的阿拉伯文名称传入欧洲后，“还原”（al-jabr）在 12 世纪被译为拉丁文 algebra，而书名的第二个字被省略。到 14 世纪，这门学科在欧洲被正式称为“algebra”。1859 年，清代数学家李善兰与英国传教士伟烈亚力合作翻译时，创造性地把英文 algebra 译为代数学。“代数”这个中文名词，让人联想到这门学科的一大特征是用字母表示数，既简洁又形象，一直沿用至今。所以数学史家克莱因曾说：在代数学方面，阿拉伯人的第一个贡献是提供了这门学科的名称。

花拉子密的《代数学》在约 1140 年被译成拉丁文，后作为标准的代数学教科书在欧洲使用了数百年，其内容、思想和方法相当广泛地影响过后来的数学家。正是由于花拉子密在代数学上做出了如此巨大的贡献，因而著名的数学史家波耶认为“代数学之父”这顶桂冠更应属于花拉子密。

第三节
16 世纪最壮观的数学成就

在前面两节，我们介绍了人类求解一次方程与二次方程的缓慢与平静的历史。在花拉子密给出完整的一次、二次方程根的求法后，这一问题已获解决。另外，我们也提到了人们对更高次方程的探索。比如，我们介绍了中国古代数学家不仅给出三次方程，还给出了一般 n 次方程的数值求解方法。不过，这只是探索的一个方向。另一个更重要的方向是，像一次、二次方程找到一般解法并给出了其根式解一样，寻找三次或者更高次方程的根式解。这引出了一段富有戏剧性的故事。

一元三次方程的故事

故事的序曲是从 15 世纪意大利数学家帕乔利（约 1445—1517）开始的。1494 年，他完成了 600 多页的数学百科全书《算术、几何、比和比例集成》，书中沿用花拉子密的解法和几何证明，广泛讨论了一次和二次方程。但在三次、四次方程方面，他总结了前人与自己艰辛探索的结果，最

后得出悲观的结论：“对于三次方程和四次方程，直到现在还不可能形成一般规则。”他认为在当时的数学中，求解三次、四次方程，犹如化圆为方一样困难。这声对以前失败的悲叹，却成为 16 世纪意大利数学家迎接挑战的号角。

故事的第一个出场人物名叫费罗（约 1465—1526），他是当时“欧洲最好的大学之一”博洛尼亚大学的数学教授。在帕乔利做出悲观结论后不久，大约在 1515 年，费罗取得了一项突破性结果。他成功得到了 $x^3+mx=n$（系数为正）这类缺项三次方程的求解公式。在求解三次方程的道路上，这是一个不小的成功。但出乎现在人们意料的是，他并没有马上发表自己的开创性成果，而是对自己的解法绝对保密。这在“不发表即发霉”的今天，真是不可思议之事！在当时却有其原因。那时的社会热衷于一种公开进行的数学智力对抗竞赛。根据 19 世纪一位数学史家的描述，数学家对这样的智力对抗感兴趣，因为其结果不仅决定他们在城市或大学里的声誉，而且会决定其任期长短和月薪多少。而一个重要的新发现就成了确保竞赛中处于不败之地的有力武器。于是，在这种社会风气下，人们常把所得的发现保密，向对手提出挑战，要对手解决同样的问题，从而可以在公开竞赛中获胜，以此保住自己的职位并提高自己的声望。费罗因此保守了这个秘密，直到临终前，才将这一发现秘传给其学生菲奥尔和纳夫（约 1500—1558），后者后来成为他的女婿，并在费罗死后保存了其研究手稿。

于是，获得费罗“杀手锏”的菲奥尔在我们的故事中作为第二个人物露面了。菲奥尔本人的数学才能并不突出，他将费罗的发现视为可利用之物，也没有立即发表，而是决定保存这件秘密武器并等待时机——一个能够让他成名的机会。天遂人愿，在他面前出现了一个可以挑战的对象。

在我们这个故事中出场的第三个人物叫塔塔里亚（约 1499—1557），是意大利数学家、力学家、军事科学家。塔塔里亚原名叫丰塔纳，1499 年出生在布雷西亚的一个贫困家庭。1512 年，法国占领了布雷西亚。法国士兵在那里大开杀戒，多数居民都到教堂寻求庇护。不幸的小丰塔纳被砍了几刀，头骨三处破裂，下巴和上腭也被劈开。尽

管最后他活过来了，但上腭的伤使他留下了语言障碍，从此有了“塔塔里亚”的绰号，意大利语就是“口吃者”的意思，后来他以此绰号闻名于世。



塔塔里亚

塔塔里亚早年的不幸还不仅如此。因为家里贫穷，他无法接受正常的教育。他母亲好不容易攒了一点钱把他送到学校，但这点钱只够他在校 15 天。在学到字母表的 k 时，钱用光了，他偷了本书，后来就靠它自学读写。据说，因为没钱买纸，他只好拿公墓的墓碑来做写字板。然而，悲惨的遭遇与穷苦的生活没有湮灭这个孩子的才能，这位顽强的少年通过刻苦自学终于成才。后来，他作为教师辗转在意大利不同城市教授科学和数学。1530 年，他的一个同事给他出了两道难题。这两道题是：求一个数，其立方加上平方的 3 倍等于 5；求三个数，其中第二个数比第一个数大 2，第三个数又比第二个数大 2，它们的积是 1000。显然，这两道题相当于分别解三次方程 $x^3+3x^2=5$ 与 $x^3+6x^2+8x=1000$。

经过不懈努力后，塔塔里亚成功解决了 $x^3+3x^2=5$。这意味着，他已掌握了 $x^3+mx^2=n$（系数为正数）这类没有一次项的三次方程的求解方法。塔塔里亚会解三次方程的消息很快就传开，并传到了菲奥尔耳朵里。菲奥尔终于等到了自己一直等候的挑战者，不久他就向塔塔里亚提出了挑战。塔塔里亚起而应战。于是，我们故事中的两位人物碰面了。

两人约定于1535年2月22日在意大利米兰公开举行辩论，双方各出30个三次方程的问题，谁解出的题目多谁就获胜。菲奥尔所提交的30个问题全部属于$x^3+mx=n$类型的方程。如其中一个问题是：一块蓝宝石卖了500金币，所得利润是其成本的立方根，求其利润是多少。列方程可得$x^3+x=500$。他希望凭借自己对这一种情形的知识而取胜。塔塔里亚提交的问题不但涉及$x^3+mx^2=n$类型的三次方程，而且还需要其他数学知识。

赛期渐近，塔塔里亚有些紧张，他并没有找到不含二次项的三次方程的一般解法。在经过夜以继日的苦思冥想后，他终于在临赛前几天的1535年2月12日发现了解法。在正式辩论中，塔塔里亚只用了两个小时就轻松解出了对方的所有题目，而菲奥尔却没有解答出塔塔里亚提的任何问题。这样，塔塔里亚以30∶0的战绩大获全胜。按当时的规矩，输者需要宴请胜者及其朋友，不过，塔塔里亚高姿态谢绝了这种奖赏，仅接受了胜利者的称号。这次辉煌的胜利为塔塔里亚带来了轰动一时的声誉，同时也意味着菲奥尔在我们的故事中以不体面的方式先行退场了。

塔塔里亚为这次胜利所激励，更加热心于研究一般三次方程的解法。到1541年，他又找到了$x^3=mx^2+n$，$x^3+n=mx^2$等类型三次方程的解法。或许是出于与费罗同样的考虑，又或许是想写一本关于三次方程解法的书，塔塔里亚没有将自己的成果很快发表。于是，风波骤起，本应进入尾声的故事，又因一个重要人物的出场而被引向完全不同的方向。

这位半路杀出来的“程咬金”叫卡尔丹（1501—1576），可算是数学史中最奇特的人物了。1501年，卡尔丹出生在意大利西部的帕维亚，是一个法官的私生子。他有着丰富多彩但曲折坎坷的一生：他的本行是医生，于1526年获医学博士学位，但因私生子的身份不被认可，直到凭借高超的医术赢得很大声望后，才于1539年被米兰医学协会接纳，后成为闻名欧洲的名医；他是一位很投入的占星家，甚至画出了耶稣的星位图，这导致他在1570年被指控信奉异端邪说而入狱；他嗜赌成性，但也对这种游戏进

行了细致的科学研究，并完成了第一本研究概率的书《论碰运气游戏》(1663 年)；与他患难与共的妻子 31 岁时早逝；他疼爱的长子在陷入一桩不幸的婚姻后毒死了妻子，被处绞刑；他的小儿子生活放荡，因偷钱付赌资，被他告发而关进监狱。最终，卡尔丹以离奇的方式告别了这个世界：迷信占星术的他预测自己将死于 1576 年 9 月 21 日，为了实现自己的预言，据说他在那一天自杀了。



卡尔丹

作为科学史上的奇人，卡尔丹被誉为百科全书式的学者，在各种知识领域里展现天赋。他一生写下各种类型的著作 200 余种，广泛涉及从科学到其他领域的众多主题。在他去世后一百年，伟大的莱布尼茨概括了他的一生："卡尔丹是一个有许多缺点的伟人。没有这些缺点，他将举世无双。"

在塔塔里亚与菲尔奥竞赛后不久，卡尔丹听说了这一消息。在此之前，他对三次方程求解问题已进行过长时间的研究，却毫无结果。可以想象，他是多么渴望知道塔塔里亚这位解三次方程大师的奇妙技巧。卡尔丹曾多次写信给塔塔里亚请求告知其解法。塔塔里亚最初都拒绝了。后来，卡尔丹向塔塔里亚承诺将把他及他关于火炮方面的发现引见到米兰宫廷，塔塔里亚有些心动了。1539 年 3 月他来到米兰与卡尔丹相

见。在卡尔丹立誓绝不公开发表塔塔里亚的发现之后，3 月 25 日塔塔里亚把为记忆三种形式的三次方程求解法则而编的 25 行诗传给了卡尔丹。下面是解释 $x^3 + px = q$ 的一段诗文：

立方共诸物，和要写右边，

巧设两个数，差值同右和；

此法要牢记，再定两数积：

诸物三（分）之一，还把立方计；

既知差与积，两数算容易，

复求立方根，相减题答毕。

卡尔丹在其后不久完成的一部数学书中遵守了自己的承诺，没有发表塔塔里亚的成果。随后，卡尔丹开始更深入地钻研这个问题，经过几年的努力，他终于得到各种类型三次方程的解法及其几何证明。这期间，卡尔丹曾得到他的一个学生的帮助。这个学生的名字叫费拉里（1522—1565），是我们这段故事最后出场的人物。

费拉里 14 岁时成为卡尔丹的家仆。卡尔丹发现了他的出众才能，收他为学生和助手。费拉里 18 岁时接替卡尔丹在米兰讲学，他最大的贡献是发现了四次方程的一般解法。

费拉里一直参与卡尔丹有关三次方程求解的工作。经过合作研究，两人获得了众多进展。但是卡尔丹和费拉里发现的解法都建立在塔塔里亚的解法基础上。由于卡尔丹立下的誓言，塔塔里亚不公布其解法，他们的新发现就不能公布。后来，卡尔丹听说费罗的女婿纳夫手中持有费罗的原稿。于是，1544 年他与费拉里一起到博洛尼亚拜访纳夫，获准阅读费罗的手稿，发现费罗先于塔塔里亚得到了相同的三次方程解法。既然这种技巧已经出现在费罗的论文中，卡尔丹觉得已经没有必要再受自己誓言的约束。1545 年，卡尔丹出版了《大术》一书，将三次方程解法公之于众。书中，卡尔丹

对不完全三次方程解法的来源做了明确说明："费罗大约于30年前发现了这一法则并传授给菲奥尔，后者曾与宣称自己也发现该法则的布雷西亚的塔塔里亚辩论。塔塔里亚在我的恳求下将方法告诉了我，但没有证明。在这种帮助下，我克服了很大困难找到了证明，现陈述如下……"

虽然，卡尔丹特别提到了塔塔里亚的贡献，然而事实是，随着《大术》一书的出版，卡尔丹立刻赢得了极大赞誉，而塔塔里亚的贡献则完全被忽视了。塔塔里亚坚信本该自己享有的荣誉被卡尔丹抢去了，因此，卡尔丹的"失信"行为激起塔塔里亚的狂怒是完全可以理解的。1546年，塔塔里亚在出版的一本书中责骂了卡尔丹的背信行为，于是一场争吵不可避免地发生了。虽然塔塔里亚要求与如日中天的卡尔丹直接交锋，但当时正处于丧妻之痛的卡尔丹保持了沉默。起而应战的是费拉里，这位以脾气暴躁著称且又忠诚的学生要报答老师的知遇之恩（17岁曾在一次争执中失去右手手指）。一时间，充满火药味的信件在塔塔里亚与费拉里之间飞来飞去。1548年8月10日，在米兰的公开辩论使这场冲突达到了白热化。当天辩论时，米兰总督做评判人，米兰人蜂拥而至。辩论开始，塔塔里亚批驳费拉里解答中的错误，费拉里则指责对方不能解四次方程。双方的争论从上午十点持续到晚饭时间，结果却是不了了之。第二天，客场作战的塔塔里亚未等裁决结果就离开了米兰。费拉里宣称自己赢得了胜利。此后，塔塔里亚失去了薪金丰厚的讲学职位，抱恨而终。而费拉里却因此名声大震，并平步青云。然而乐极生悲的是，1565年，年仅43岁、已成为大富翁的费拉里忽然去世，据说是他妹妹图谋他的财产，毒死了他。

这些就是围绕着三次方程求解所发生的故事，激烈、复杂、有趣却又不免有点荒唐。由于卡尔丹最早发表了求解三次方程的方法，因而数学上三次方程的解法被称为"卡尔丹公式"，塔塔里亚之名湮没无闻。这对塔塔里亚来说似乎欠公平。不过，在历史上，这类争夺优先权的论战又何止这一桩呢？随着时间的推移，当时对于个人如此重要的事，对后人而言却不过是"古今多少事，都付笑谈中"而已。

抛开个人恩怨得失，我们看到的是，在这出悲喜剧谢幕时，数学家们已经知道了怎样解三次和四次方程。

16世纪最壮观的数学成就

我们先来考察一下上面提到的缺二次项的三次方程 $x^3+mx=n$（系数 m、n 为正数）的解法。对此类方程，费罗与塔塔里亚都得到了解法，但均未发表。卡尔丹在《大术》第十一章以 $x^3+6x=20$ 为例，给出了这类方程的解法。用现代符号，我们大致阐述一下其思路。

考虑恒等式 $(a-b)^3+3ab(a-b)=a^3-b^3$，将这一恒等式与 $x^3+mx=n$ 对照，我们可以设 $x=a-b$，这只需 $m=3ab,\ n=a^3-b^3$ 即可。这样，只要我们选取适当的 a、b，则 x 由 $a-b$ 给出。于是问题转化为求满足 $m=3ab,\ n=a^3-b^3$ 这一条件的 a、b。

这一条件又可以变化为：$a^3(-b)^3=-\dfrac{m^3}{27},\ a^3+(-b)^3=n$。

这意味着 a^3、$-b^3$ 是二次方程 $x^2-nx-\dfrac{m^3}{27}=0$ 的两个根。于是原来的三次方程求解问题转化成了当时人们已熟知的二次方程。套用二次方程的求根公式即可得到

$a^3=\dfrac{n}{2}+\sqrt{\dfrac{n^2}{4}+\dfrac{m^3}{27}},\ b^3=-\dfrac{n}{2}+\sqrt{\dfrac{n^2}{4}+\dfrac{m^3}{27}}$。

最终得出，$x=a-b=\sqrt[3]{\dfrac{n}{2}+\sqrt{\dfrac{n^2}{4}+\dfrac{m^3}{27}}}-\sqrt[3]{-\dfrac{n}{2}+\sqrt{\dfrac{n^2}{4}+\dfrac{m^3}{27}}}$。

这一由符号表示的结果与卡尔丹的结论是相同的，只是他的文字叙述要复杂得多：

“将一次项系数的$\frac{1}{3}$立方；加上常数项之半的平方；并取整个的平方根。现在对这个数分别加上及减去常数项之半，那么，第一个数的立方根与第二个数的立方根之差便是未知数的值。”套用这一公式，$x^3+6x=20$解为：$x=\sqrt[3]{\sqrt{108}+10}-\sqrt[3]{\sqrt{108}-10}$。

为了符合当时的严格标准，即任何代数结果必须经过几何方法的证明才被确认为真，卡尔丹用几何方法给出了上述方程解法的复杂证明，这在当时被认为是极其重要的一步。与二次方程证明中使用的正方形填补法相类似，卡尔丹的证明本质上是一种“立方体填补法”。

由于当时的数学家一般不接受负数，于是，卡尔丹不得不在后面的两章继续研究$x^3=mx+n$，$x^3+n=mx$的类型。卡尔丹在书中第十四章至第十六章探讨了三种缺一次项的三次方程，在第十七章至第二十三章探讨了7种四项俱全的三次方程。也就是说，卡尔丹根据二次项、一次项和常数项的位置不同，把三次方程分为13种类型，并用13章的篇幅对每一种类型给出了不同的求解法则及证明。而现在允许系数为非负数，因此对三次方程只需要讨论一种统一的形式，即$ax^3+bx^2+cx+d=0$，其中$a\neq 0$。

如何解决这类不缺项的三次方程？谜底揭开后倒是很简单的，需要的只是一个代换而已。事实上，令$y=x+\frac{b}{3a}$，即可消去二次项，将其转换为已解决的$y^3+py+q=0$类型的方程。

《大术》另一重要成果是给出了由费拉里发现的四次方程解法。它出现在《大术》第三十九章中：“还有另外一个法则，并且，比前一个法则更为壮观。这就是费拉里提出的法则，他应我的要求，将其发现交给我，根据费拉里法则，我们可以求出所有四次方程的解。”我们用现在的符号大致说明一下费拉里的思路。

先看一下如何解缺三次项的四次方程$x^4+bx^2+cx+d=0$。

对前两项配方得，$\left(x^2+\frac{b}{2}\right)^2=-cx+\frac{b^2}{4}-d$。

费拉里的基本思路是通过引入参数把等式的两边都配成完全平方形式。

为此引入参数y，方程两边同加上$2y\left(x^2+\frac{b}{2}\right)+y^2$，可得到

$$\left(x^2+\frac{b}{2}\right)^2+2y\left(x^2+\frac{b}{2}\right)+y^2=2y\left(x^2+\frac{b}{2}\right)+y^2-cx+\frac{b^2}{4}-d$$

$$即\left(x^2+\frac{b}{2}+y\right)^2=2yx^2-cx+\left(yb+y^2+\frac{b^2}{4}-d\right)\quad (*)$$

左边已经是完全平方式，右边为了成为完全平方式，只需右边多项式的判别式为0，即$c^2-4\cdot 2y\cdot\left(yb+y^2+\frac{b^2}{4}-d\right)=0$。整理此式，可以看到这是一个关于$y$的三次方程。根据卡尔丹公式，一定可以解出$y$。然后把$y$代入（*）式，（*）式两边都是完全平方式。因此通过两边开方，得到关于x的两个二次方程，于是可求得x。

那么一般的四次方程$ax^4+bx^3+cx^2+dx+e=0$又如何求解呢？可作代换$y=x+\frac{b}{4a}$，于是原方程将转化为已获解决的缺三次项的四次方程$y^4+my^2+ny+k=0$。

这就是费拉里给出的四次方程的大概解法。其方法有普遍意义，并成为现代代数方程理论中的标准解法。但正如我们已经提到的，因为当时受限于系数为正，所以《大术》一书不得不考虑复杂得多的情形。事实上，卡尔丹在书中列举了20种不同类型的四次方程，除这些“事实上是最一般”的以外，卡尔丹指出“还有67种其他类型的四次方程”！即便对列出的20种类型的四次方程，卡尔丹也没有再像处理三次方程那样展开，而只是概述了基本过程，并选出部分典型题目举例说明。在书的结尾他写道：“用五年时间写就的这本书，也许可以持续几千年。”

卡尔丹的自夸是有道理的。他的《大术》一书使解方程的艺术达到了新高度，当初认为三次方程不能解的观点被彻底粉碎。书中关于三次方程和四次方程的解决被数学史家伊夫斯称作“16世纪最壮观的数学成就”。此外，作为代数发展史上里程碑的《大术》还推动了方程理论的探讨。卡尔丹在书中详细讨论了方程的负数根，成为欧洲第一个允许二次方程和三次方程负根存在的人。他指出一个三次方程能有三个根，一个四次方程能有四个根，这些工作被认为是代数方程理论的早期成果。

从整个数学的发展来看，这本16世纪最重要的数学著作更具有承前启后的重要意义。在承前方面，三次、四次方程求解上承二次方程千余年的求解成果。《大术》在许多方面保持着花拉子密的风格：系数只能为具体的数；要求方程的系数为正数，所以不得不对方程分类处理；没有使用代数符号，对方程、方程的解法与规则的说明都用冗长的语言叙述；使用几何证明等。这并非卡尔丹的局限，而是那个时代的局限。在启后方面，他首先注意到三次方程求解的不可约情形，并由此提出虚数问题，这对方程理论研究有重要意义。更重要的是，《大术》对三次、四次方程的解决成为17世纪和18世纪数学家对更高次方程展开一系列漫长而影响深远的探讨的起跑点。

第四节
另两位代数学之父

在第二节中，我们介绍了两位代数学之父。一位是丢番图，赢得这一称号主要是因为他引入了未知数，创设了未知数的符号，并有建立方程的思想。另一位是花拉子密，他主要凭借系统论述了求解一次、二次方程的方法而享有这一美誉。除这两人外，历史上还有两位数学家被冠以“代数学之父”，他们是韦达与高斯。我们在这一节就来介绍这两位及其在代数学方面的贡献。

韦达与符号代数

韦达（1540—1603），是16世纪最伟大的数学家。韦达不是专职数学家，他本行是律师，并从事政治活动，曾以律师身份在法国议会里工作。但他几乎把所有的空闲时间都用在了数学研究上。他专注数学到什么程度呢？有时解决某些问题，可以连续几夜不睡。



韦达

关于韦达有一些迷人的趣事。比利时一个外交官向法国国王亨利四世夸耀说，法国没有一个数学家能解他们国家罗曼纽斯（1561—1615）在 1593 年提出的问题，它需要求解一个 45 次方程。国王召来韦达，让他看了方程。韦达发现它关联着三角函数，几分钟就给出了两个根，接着解出了另外 21 个根，忽略了负根。之后，韦达挑战罗曼纽斯，让他解阿波罗尼奥斯的问题（画一个圆与三个给定的圆相切），但罗曼纽斯没能用欧几里得的方法得到解。当他看了韦达的绝妙解法后，就去拜访韦达，两人发展出一段真诚的友谊。

另一个故事是，在法国与西班牙的战争中，韦达成功破译了西班牙的几百字的密码，使法国在两年的对西战争中赢得了先机。然而，西班牙国王菲利普二世根本不信有人能破解那个密码，于是他向教皇抱怨说法国人用魔法对抗他的国家。

韦达一生写了大量三角、代数和几何的书，多数都是自费印刷发行的。他最主要的贡献表现在代数方面。

韦达最为我们所熟悉的成就是他关于根与系数关系的发现，事实上，一般人知道韦达的名字正是由于这一结论被命名为“韦达定理”。韦达定理在 1591 年完成，在韦

达去世后才出版的一本书中公布：二次方程中，如果第二项的系数是两数和的相反数，第三项的系数是这两数的乘积，那么这两个数就是这个方程的根。用我们所熟悉的现代形式表示就是，设x_1，x_2是二次方程$x^2+px+q=0$的两个根，则$x_1+x_2=-p$，$x_1x_2=q$。

后来，人们把韦达定理推广到更高次的方程。一般情形的一元n次方程的根与系数的关系，是由荷兰数学家吉拉德于1629年提出的，其证明则是在笛卡儿于1637年提出因式定理、高斯于1799年证明了代数基本定理以后才完成的。

在方程论方面，韦达的另一成就是改进了三次和四次方程的解法。如缺项三次方程$x^3+3ax=b$，韦达解法的思路如下：令$x=\frac{a}{y}-y$，则方程变为$\left(\frac{a}{y}-y\right)^3+3a\left(\frac{a}{y}-y\right)=b$。经整理后，得出$y^6+by^3-a^3=0$，这个以$y^3$为未知量的二次方程可解。求得$y^3$后，再求$y$与$x$即可。

韦达在代数学方面更重要的贡献，则是他在丢番图的启发下引入了符号代数思想。

1591年，韦达发表其代表作《分析术引论》。与以往曾有人偶然地用到字母与符号不同，韦达是第一个有意识且系统使用了字母与符号的人。由此，他创立了一般的符号代数。在他手中，字母不仅用于表示未知量或未知量的幂，而且用以表示一般的系数。

在前面的介绍中，我们看到从古埃及和古巴比伦时代起，数学家们仅解决带有数字系数的一次、二次、三次和四次方程。因而，像$3x^2+5x+6=0$和$4x^2+6x+7=0$这两个方程被认为是不一样的，尽管很明显地存在着一种同样的方法来解决这两个方程。此外，为了避免负数，在很长的一段时期内，人们将诸如$x^2-7x+8=0$这样的方程用$x^2+8=7x$的形式来处理，这样，同次方程中就有许多种类型并且每一种都要分别求解。而当韦达首先引入字母系数（并接受负系数）后，所有的二次方程便可以统一写成$ax^2+bx+c=0$的形式并用统一的方式来处理。在这里，a、b、c这些字母系数可以

表示任意数，x 则代表未知量。

韦达将新型的代数叫作“类的运算”，以区别于数字计算。他清楚，当研究一般二次方程 $ax^2+bx+c=0$ 时，他所处理的是完整一类的表达式。在书中，他还划分了算术与代数的界限。在他看来，代数是作用于事物的类别或形式上的方法，是类的运算。而算术和数字系数的方程则是与数打交道，是数字计算。通过引入类的运算，韦达在代数学的发展中迈出了一大步，他使代数成了研究一般的类和方程的学问。

这种处理的明显优点是我们所熟悉的：如果证明了某种求解 $ax^2+bx+c=0$ 的方法是正确的，那么，这种方法就可以确保求解无穷多个具体方程的正确性。这就使代数证明的普遍性成为可能。

不过，韦达的代数符号与现代符号仍相去甚远，其书中所记载的方程形式和现在的还有很大差别。对于习惯于现代数学的读者来说，他选用的符号并不优良，他用辅音字母表示已知量，元音字母表示未知量的方法没有沿用下来。用 a、b、c 表示已知量，x、y、z 表示未知量的习惯用法是法国数学家笛卡儿继韦达之后提出的。另外，韦达书中还附有过多的文字说明（如相等、相乘等没有专用符号，而是用文辞表示）。举一个例子，方程 $x^3+30x^2+44x=1560$，在韦达的书中表示为：1C+30Q+44N，aequatur 1560。在韦达之后，数学符号继续被引入，符号体系不断得到改进和完善，最终形成了我们现在所熟悉的代数符号体系，其中 17~18 世纪的笛卡儿、莱布尼茨和欧拉等都在这方面做出过重要的贡献。

尽管如此，韦达的确是朝着用字母表示方程的方向迈出了非常重要的一步，而他在建立代数符号体系方面所做出的卓越贡献彻底改变了代数学的面貌，因此他被称为“代数学之父”。如果认为代数学本质上最大的变革是从符号体系方面开始的，那么作为符号代数奠基者的韦达获得这一称号是名副其实的。

高斯与代数基本定理

一个代数方程是否总有解？如果把数系的范围限制在整数、有理数甚至实数内，我们无法保证这一点。最简单的例子是，$x^2+1=0$ 在整数、有理数甚至实数范围为无解。那么，在复数范围内考虑，情况又将如何呢？许多人猜测答案是肯定的。1746 年，达朗贝尔给出这一结果的一个不太严格的证明。1799 年，22 岁的高斯在其博士论文中第一个较严格地证明了：实系数的 n 次方程总是包含至少一个复根。当然，在迈出这困难的一步后，人们进而较容易地推出 n 次方程有且仅有 n 个复根（重根按重数计算）。

在高斯的证明之后，柯西等数学家又给出了若干证明。而充分认识到这一结果重要性的高斯本人在 1815 年、1816 年给出了另外两个证明。最初人们的证明都假定系数是实数，即任意实系数多项式方程至少存在一个复根。1849 年在庆祝取得博士学位 50 周年的纪念会上，高斯发表了这一结论的第四个证明。这一证明首次把方程的系数推广到复数，即高斯证明了：任意复系数多项式方程至少存在一个复根。容易知道，正系数的方程，根可能不是正数。整系数的方程，根可能不是整数。有理系数的方程，根可能不是有理数。实系数的方程，根可能不是实数。但复系数代数方程，它的根还是复数。利用代数方程，没法使复数家族增加新成员了，这叫作“复数系统的代数封闭性”。

“一元 n 次方程至少有一个复根”这一在方程论中具有中心地位的结论，被称为**代数基本定理**。这个名称源于当时方程理论和代数学基本就是一回事，而这一定理保证了代数方程根的存在性。

代数基本定理的获证在代数学发展史上具有里程碑的意义。利用这一结果，人们可以很快推出其他结果，如一个 n 次多项式能分解成线性因式的乘积，任何实系数多项式可表示成一次和二次实系数因式的乘积等。而高斯正因为在证明这一定理中做出了关键性贡献，所以有数学史家也把高斯称为“代数学之父”。

代数基本定理的证明，产生了多方面的影响。

一方面，这一定理的成立必须依赖于数系的扩充。数系的扩充是贯穿于数学历史的一条明显的红线。但与人们通常的想象不同，数系的扩充经历了极其漫长、复杂、曲折的过程。我们以负数、复数的引入为例简单说明一下。

古代中国，最晚到公元 1 世纪下半叶，已经有了明确的负数概念。但在西方，负数概念被认可经历了极长的时间。比如，17 世纪的法国著名数学家帕斯卡仍认为：“从 0 减去 4 纯粹是胡说。”

但当面对负数能否作为方程的系数与方程的根这一问题时，东西方态度则非常一致，他们都持小心翼翼的否定态度。

在方程系数方面，我们现在对其正负不作区分。比如说，二次方程的一般形式我们可以记作 $ax^2+bx+c=0$，其中的系数 a、b、c 可正可负，就是说我们对系数的正负是一视同仁的。然而在古代，东西方迈出这一步都非常不容易。我们已经提到，花拉子密和卡尔丹在解二次、三次方程时，都限定系数为正。这一点到代数学之父韦达时，尚无改变，他拒绝用字母系数表示负数。直到 17 世纪中叶，西方才有数学家允许字母系数既可以代表正数，又可以代表负数。在此以后，西方人才慢慢接受了负系数。在我国，直到宋元时期，才逐步在方程中引入了负系数。

在方程的根方面，我们现在既取正根，也取负根。但在数学发展史上，迈出这一步同样经历了漫长的岁月。如西方历史上第一个引入负数运算的丢番图，只接受正有理根，负根与无理根都被他忽略。之后的长时期内，多数西方数学家都拒绝接

受方程的负根。比如，韦达在解方程碰到负根时就把它舍去。在我国，从产生负数的公元 1 世纪到数学最鼎盛的宋元时期，在对待方程的负数根方面采取的态度也都是不予考虑。

随着数学家最终接受负数，负数在数学中的重要性与必要性得以体现。在前面章节，我们已经看到花拉子密、卡尔丹在处理方程时，因为没有负系数的思想而不得不把我们现在看来相同的方程分为不同的类型，从而带来许多不必要的复杂，并使方程求解变得烦琐不堪。通过古今这种繁简的强烈对比，我们应能从中领略到负数引入的必要性。此外，引入负数后，许多无解的方程也有了解。

下面再来简单看一下复数的引入。

如我们所熟悉的，在解二次方程时就会遇到负数开平方问题。然而，复数的引入却是解三次方程的产物。事实上，人们早期遇到 $x^2+1=0$ 之类的方程时，采用的方式是回避，即认为这样的方程无解。事情的转折来自卡尔丹三次方程求根公式的发表。比如我们考虑一下三次方程 $x^3=15x+4$ 。利用卡尔丹公式可以得到 $x=\sqrt[3]{2+\sqrt{-121}}+\sqrt[3]{2-\sqrt{-121}}$ 。要知道 16 世纪数学家对负数尚持怀疑态度，公式中出现负数的平方根在当时的数学家看来绝对是荒谬与不可接受的。那么，能否使用曾有效的回避办法，将这一方程当作不可解的三次方程而予以排除呢？就像人们不承认负数，就避开 $x+4=2$ 之类的问题一样。然而，这一次事情似乎更麻烦。因为对上面这个三次方程来说，恰恰可以很容易验证出它有三个不同的完美实数根：4 及另外两个实根 $-2\pm\sqrt{3}$ 。所以，现在的困难在于，我们没有理由抛弃上述的方程，即认为它没有解或者不值得解，而是应该去寻找一种解释，说明为什么我们的根式解与真正解之间存在着似乎不可调和的矛盾。后来，人们认识到，在用求根公式解三次方程时不可避免要陷入一类新数的困惑之中。特别是，如上面例子所见到的，当三次方程的三个根是互不相同的实数时，用卡尔丹公式得到的结果是由两部分都含有根号下负的数组成，疑难无可避免。卡尔丹在书中指出了这类疑难的存在并进行了一些研究但并未解

决。后来，意大利数学家邦贝利迈出了勇敢的一步。在1572年的论文中，他接受一类新的数即我们现在所称的“虚数”，通过熟练应用这类虚数，他发现可以得到方程$x^3=15x+4$的实数解。于是，虚数成为运载数学家从实数三次方程到达其实数解的必要工具。也就是说，人们从熟悉的实数领域出发并最终回到实数，但中途须进入一个当时人们不熟悉的虚数世界以完成这一旅程。这种处理方式对当时数学家来说有些不可思议，然而却可以在形式上有效地解决用卡尔丹公式解三次方程出现的矛盾。于是，邦贝利的新思想及由之产生的新方法，使人们在一定程度上开始认真看待复数。复数作为一种有用的工具，被数学家引入了，此后它被数学家们越来越熟练地使用。然而，即使到了数学家已经对复数的运算达到相当熟练的程度，复数作为数的地位仍然无法得到确立。

代数基本定理的认识与证明为普遍认可复数提供了新的理由与推动力。正如我们已经看到的，历史上容许哪些根不容许哪些根一直是有争议的。在16、17世纪，甚至到19世纪，许多数学家不仅反对虚根，连负根也不接纳。如此一来，同样是二次方程，就可能出现二个根、一个根或者没有根等多种情况。这种偏狭的见解最终还是被宽容的意见所取代，即把复根都包括在内。在这种宽容的观念下，产生了美妙的结论：“一元n次方程有且仅有n个根（重根按重数计算）。”这一漂亮的结果使数学形成一种完美的理论，但它必须接受复数的存在，即结论是建立在复数基础上的。于是，代数基本定理的证明巩固了复数的地位。

代数基本定理的证明产生的另一个影响，体现了它是一种存在性证明。在此之前，人们更习惯于一种构造性证明。换句话说，以前的存在性大多是通过实际获得或显示出问题中的量而建立起来的。例如，二次方程解的存在性，是通过求根公式把满足方程的解显示出来。而高斯探讨代数基本定理的方法与之迥然不同，他开创了探讨数学中整个存在性问题的新途径。这种证明告诉世人宝藏的存在，但并未说明其藏宝地点。因此，根据已经证明的代数基本定理，人们只知道方程的解一定是

存在的。然而，这些存在的解是否都能用公式表示出来，高斯等人的证明没有对此透露任何信息。

当然，如上面已经介绍的，到 16 世纪时，数学家们已经能够对五次以下的方程给出根式解。于是剩下的问题是，五次或更高次方程是否也存在这样的根式解。对这一问题的思考自 16 世纪起困惑了数学界 200 多年，直到 19 世纪才被彻底解决。这正是我们下一节要探讨的问题。

第五节
两颗璀璨的数学流星

在卡尔丹《大术》发表后，破解五次方程之谜成了数学中最迷人的挑战之一。然而，两个多世纪中，寻找五次方程根式解的努力全以失败告终。作为努力的成果之一，人们找到一种数学变换，可以把一般五次方程简化为简单的 $x^5+px+q=0$ 的形式。不幸的是，这种似乎更易处理的形式仍然是难以逾越的障碍。“四周的城墙已经增高，问题仍躲在最后的城堡里，绝望地捍卫着自己。谁会成为迫使它投降的幸运之人?”19 世纪，数学史上两位非凡的数学天才在数学天空中飞过，他们留下的数学遗产终于给出了问题的完美解答。

序　　幕

1770 年，被拿破仑誉为“数学科学高耸的金字塔”的法国数学大师拉格朗日（1736—1813）发表其杰出论文《关于方程的代数解法的思考》，为陷于困境中的这一研究开辟了一条新路。

在文章中，拉格朗日致力于寻找对二次、三次、四次方程普遍适用的根式解法。他想将这种解法推广到五次方程以解决问题。拉格朗日几经努力终于发现：一个已知方程的根，可由另外一个辅助方程的根的对称函数来表示。这个辅助方程被拉格朗日称为**预解式**。他发现利用预解式，二次、三次与四次方程都可以实现降次，从而它们的求解问题可顺利解决。但当完全相同的过程运用到五次方程时，意外发生了。在低于五次方程中发挥漂亮作用的方法在五次方程上遭到彻底失败。拉格朗日由此总结道："因此，用这些方法推导出五次方程的解——最著名的和重要的代数问题之一——是不可能的。"

虽然拉格朗日未能获得成功，但其意义深远的尝试已把一种激动人心的新思想引入了研究。特别是，他引入了置换的新思想，并指出了方程的解的置换特点可能与方程是否有公式解有关。后来的研究证实了拉格朗日的预见，人们发现根的置换理论确实是解代数方程的关键所在，是"整个问题的真正哲学"。

下一个在冲击这一问题时做出贡献的是意大利数学家鲁菲尼（1756—1822）。1799年，在两卷厚达516页的论文《方程的一般理论》中，鲁菲尼声称自己证明了：一般五次方程不能通过一个公式解出。然而，他的证明极其复杂。后来，他又多次改进自己的证明。1813年，他发表了一篇最终证明，打算给出更严格和不怎么深奥的证明。

然而，不幸的是，鲁菲尼的证明看来鲜为人知，而且未被普遍接受。他曾几次寄论文给拉格朗日，但未得到回复。数学界对他曲折而复杂的证明持怀疑态度，当他于1822年去世后，他的工作几乎被遗忘了。

高斯在发表代数基本定理第一个证明的同一年，曾表达过他对五次方程根式解的看法："在许多几何学者的努力后得到一般方程代数解的希望仍然渺茫，似乎越来越可能这种解是不存在和矛盾的。"接着，他加了一句让人感兴趣的注解：严格证明五次方程不可能有公式解也许不是那么困难。关于这个主题，高斯再也没有发表过其他言论，也没有给出这个"不那么困难的"证明。

于是，到19世纪初，在经过众多数学家的顽强努力后，用根式解高于四次方程的问题仍悬而未决，“它好像在向人类的智慧挑战”。接过攻坚的接力棒，迎接挑战的是两位年轻的数学天才：阿贝尔与伽罗瓦。

阿贝尔：天才与贫困

1802年8月5日，阿贝尔出生在挪威一个小村庄里，父亲是村子里的穷牧师。13岁时，阿贝尔进入奥斯陆一所教会学校学习。15岁时，他遇到了他一生中最幸运的事：霍尔姆伯成为他的老师。这位本人缺乏创造力，却知道并能鉴别数学杰作的数学家发现了阿贝尔的数学天赋。在他的热心指导下，阿贝尔令人惊讶的数学才能开始得到展现。霍尔姆伯后来回忆道：“阿贝尔沉迷数学，他以惊人的热忱和速度向这门科学进军。在短期内他学了大部分的初级数学，在他的要求下，我私人教授他高等数学。过了不久，他自己读法国数学家泊松的作品，看德国数学家高斯的书，特别是拉格朗日的书。他已经开始研究几门数学分支。”就这样，阿贝尔很快掌握了经典著作中最难懂的部分，包括高斯的《算术研究》。从那以后，真正的数学就不仅是他的严肃工作，而且成为他着迷的爱好。若干年后，有人问起他是怎样设法迅速地赶到前面去的，他回答：“靠学习大师，而不是学习他们的学生。”

阿贝尔18岁的时候，父亲去世了，照顾母亲、精神不正常的哥哥和五个弟妹的家庭重担落到了阿贝尔一人身上。在艰难困苦中，他坚持不断地工作。中学最后一年，阿贝尔开始了他第一个抱负非凡的冒险：解决一般五次方程的根式求解问题。开始时

他以为自己找到了求解公式，他的老师霍尔姆伯找不出论证的破绽。后来，这篇文章被寄给一位丹麦数学家。这位丹麦数学家也没有发现论证本身的错误，但他明智地建议阿贝尔用一个例子说明自己的方法。接受建议的阿贝尔很快发现自己的推理中确实存在漏洞，这促使他几年后走到一条相反的道路上。1824 年，阿贝尔成功证明了一般五次方程及高于五次的方程不存在根式解。

1825 年，在朋友们的帮助下，阿贝尔得到政府的资助得以拜访欧洲其他国家的数学家。然而，从这些已成名的数学家那里，阿贝尔没有得到太多有益的帮助。唯一幸运的是在柏林他遇到了一位业余数学爱好者克列尔，这是他一生中第二个对他的事业有极大帮助的人。克列尔在 1826 年创办了世界上第一本专门从事数学研究的定期刊物《纯数学与应用数学杂志》。阿贝尔在该杂志前三卷上共发表了 22 篇论文。论文的内容涉及广泛，包含了方程论、无穷级数、椭圆函数论等多方面的内容。

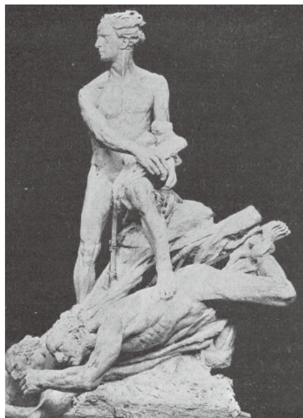
1826 年，在巴黎的阿贝尔完成了一篇非凡的论文《论一类极广泛的超越函数的一般性质》，他抱着极大期望把文章提交给法国科学院。这一“也许称得上这个世纪最伟大的数学发现”（雅可比语）、“给数学家们留下了够他们忙上五百年的东西”（埃尔米特语），被交给勒让德、柯西审阅。年老的勒让德感到论文手稿“几乎看不清楚”“简直无法辨认”，年富力强的柯西则“忙于孵自己的蛋”，根本无暇检查阿贝尔已孵出的大蛋，一篇杰作完全被忽视了。寄出论文后满怀希望等待的阿贝尔，几年中一点儿论文的消息也没听到。直到 1830 年，为了弥补过失，法国科学院授予阿贝尔（和雅可比一起）数学大奖。然而，可怜的阿贝尔已经永远无法领取这个奖了。

1827 年，由于债台高筑又未能找到合适的职位，阿贝尔返回祖国挪威。此后的生活变得更为悲惨。起初他没有找到任何固定的工作，只能以私人授课维持生计。1828 年他总算在一所大学里担任代课教师，但不久后，这个受贫穷打击的人又“穷得就像教堂里的老鼠”。一直缠绕他的贫困和伤心，把他的身体搞垮了，他得了肺结核病。1829 年 4 月 6 日，年仅 27 岁的阿贝尔走完了自己短暂的一生。

阿贝尔去世后两天，热心的克列尔寄来一封信说，他将被任命为柏林大学的数学教授。可以改变他贫困处境的机会终于来了，然而，这一切已经太迟了。一个极富数学才华的青年已经怀着对生命的眷恋、对数学的无限热爱，离开了这个世界。一颗光辉夺目、明亮无比的流星在 19 世纪的数学天空中一闪而逝。

姗姗来迟的荣誉在阿贝尔死后终于降临。为了纪念自己国家最伟大的天才数学家，挪威政府于 2002 年 8 月 5 日，阿贝尔诞辰 200 周年之际，设立了阿贝尔奖。阿贝尔也出现在一些小型张邮票上面。挪威著名雕塑家古斯达夫·维克朗（1869—1943）还于 1908 年完成了著名的阿贝尔雕塑（如下面的右图所示）。





数学则以其特有的方式为阿贝尔树立了更为长久的纪念碑。今天，任何一个学习高等数学的人都会在许多定理和理论中遇到阿贝尔的名字。作为对他多项成就的肯定，数学中以阿贝尔命名的术语、概念和定理有几十个。在其短暂的数学生涯中，他留下了多方面的非凡贡献：他求解了第一个积分方程；他为无穷级数理论奠定了严密的基础；他是椭圆函数论的奠基人之一，等等。此外，在与我们主题相关的方面，他首次成功证明了一般五次和高于五次的方程不存在根式解。这项了不起的成就，是年仅 22 岁的阿贝尔在 1824 年经过几个月的紧张工作后获得的。由于不知道鲁菲尼的工作，阿贝尔的证明非常迂回复杂。后来他又发表了两个更精简的证明。

阿贝尔没有忽略自己这一发现的重要性，他自费印刷了《论代数方程，证明一般五次方程的不可解性》的论文。但为了节约费用，阿贝尔将文章压缩为仅仅6页。然后他把自己的小册子寄给几位数学家，决定用这一证明作为自己的“名片”，认为这“将可能是我最好的介绍”。但他为自己过分的节俭付出了代价。对大多数数学家而言，这几乎像密码一样的论文太晦涩了。其中的一份小册子寄给了伟大的高斯。但高傲的高斯大概不相信这位尚名不见经传的小人物能在如此短的文章中解决他自己也未能解决的世界难题，或许他以为这是一位哗众取宠的年轻人的闹剧，于是根本没看一眼就把阿贝尔的伟大成果抛到了一边。在高斯死后，人们在他的遗物中发现阿贝尔的论文甚至没有被裁开。数学文献中最伟大的杰作之一当时竟然没有读者！在这一不幸的事件中，受损失的不单是阿贝尔，还有整个数学学科。

阿贝尔的工作，终结了人们一直探索中的代数方程根式求解问题。自意大利数学家成功找到了三次、四次方程的求根公式后，得陇望蜀的数学家们在随后两个多世纪中一直苦苦寻觅五次方程的求根公式。然而200多年里，许多著名数学家的努力都徒劳无功。但许多人仍然相信这样的公式是存在的，缺乏的只是巧妙的思路。拉格朗日在失败后怀疑这样的公式可能不存在，随后鲁菲尼又给出了一个有漏洞的复杂证明。最后阿贝尔出现了，他的证明给出了革命性的观念：人们一直寻求的东西没有找到，只是因为它们根本就不存在。他严格证明了一般的五次及高于五次的方程没有根式解。换言之，只使用加减乘除或开根号这几种代数运算，无法用一个公式表示出五次及五次以上方程的解。阿贝尔的否定性回答在数学史上是一个里程碑式的认识。

阿贝尔的工作指出，高于四次的一般方程不能用根式解出，但这并没有否定还有许多种特殊类型的高次方程确实可以用根式解出。在用根式解特殊类型的高次方程方面，高斯做出过重要贡献。他曾在1801年的《算术研究》中详尽考察了所谓的分圆方程 $x^p-1=0$（p为素数）。他证明这类方程总能归为解一串较低次的方程，从而可以用根式求解。阿贝尔也考虑了一类能用根式求解的特殊方程，现在称为阿贝尔方程。在

此过程中，他还引入了两个重要概念：“域”和给定域中的“不可约多项式”。因早逝，阿贝尔未能解决用根式求解的全部方程的问题。

于是，当阿贝尔在方程求解方面做出伟大发现后，关于用根式解方程的全部问题在新的基础上提出来了：找出一切能用根式解出的那些方程。换句话说，就是找出方程能用根式解出的充分与必要条件。这个问题最终由一位与阿贝尔有着同样天赋和同样不幸命运的传奇人物伽罗瓦彻底解决。下面且让我们先去了解一下他的一生。

伽罗瓦：天才与愚蠢

1811 年 10 月 25 日，伽罗瓦出生于巴黎。他的父母都是很有教养的人。他最初是在家接受母亲的启蒙教育，一直到 11 岁。1823 年 10 月，伽罗瓦进入一所久负盛名的中学。15 岁时，他陷入对数学如饥似渴的学习中。他的一位老师说：“他被数学迷住了心窍。”他的学业报告单也表明，他对所有别的课程都不重视，而单单专心于他新找到的这门心爱的学科。他的一位老师对此评价说：“该生只宜在数学的最高领域中工作。这个孩子完全陷于数学的狂热之中。我认为，如果他的父母允许他除了数学不再学习任何东西，将对他是最有好处的。否则，他将在这里浪费时间，并且他所做的只会使他的老师们痛苦，而他自己则被惩罚压垮。”

伽罗瓦的数学天赋不断得以展现。他读几何书就像看小说那样容易，据说勒让德的《几何原理》这本老师打算要讲两年的新教材，他只用 2 天就掌握了。而学习代数学不久，他就开始研读拉格朗日的原著了。伽罗瓦的天才与努力使他的数学水平很快

超过了他的老师。于是，他直接学习当代大师们写的最新著作。他迅速地汲取那些最复杂的思想。如同阿贝尔一样，学习和研究数学大师的经典著作，是伽罗瓦获得成功的重要途径。如同阿贝尔一样，不了解鲁菲尼与阿贝尔工作的伽罗瓦开始了一次野心勃勃的尝试。16 岁时，他花了 2 个月时间尝试解五次方程。与阿贝尔惊人相似的是，最初他也自信找到了公式，但随后他也发现了错误。这个挫折同样促使他在几年后走向一条相反的道路。

期间，伽罗瓦这个厌倦了正规学习的神童和他那愚钝的教师之间的冲突日益增长了。在学校的记事中记载着："这个学生只是怕受惩罚才完成作业，上两周他破例地做了些功课……他经常佯作雄心勃勃、与众不同，性格的怪僻又使他和同学格格不入。"

1828 年 6 月，准备不充分的伽罗瓦第一次参加巴黎综合工科学校的入学考试，这是法国最好的学校。结果"在某个领域的知识太多，而其他领域的知识太少了"的他落榜了。这次失败使他非常痛苦，无奈下他回到原来的学校继续读书。幸运的是，他遇到了一位杰出的数学教师理查德，这位老师之于伽罗瓦正如霍尔姆伯之于阿贝尔。这位数学教师意识到伽罗瓦的天才并且努力帮助他、鼓励他。

1829 年 4 月，伽罗瓦发表了第一篇论文。5 月，他向科学院提交了一篇包含基本发现的论文。不幸，这份论文因审稿人柯西的疏忽大意而丢失。很快，打击接踵而至。7 月 2 日，他的父亲，一位自由主义派的镇长因屈服于政敌的恶意攻击自杀身亡。8 月 3 日，服丧期间的他再次参加他所渴望进入的巴黎综合工科学校的入学考试。但这次考试同样注定不会有好结果。在他考试的口试阶段，主考官对一个对数级数的结果提出了质疑。伽罗瓦没有解释他的想法，而是说这个结果很显然，这就轻视了考官的智力。当发现自己的面试很糟时，据说被激怒的伽罗瓦把黑板擦扔到了主考官的脸上。很自然地，他又一次落榜了。最后他不得不考入高等师范学校。

这一连串打击与厄运，在伽罗瓦心上留下了深深的印记。但他仍狂热地继续着他的数学研究，让自己完全沉浸在数学中。1830 年，他完成并发表了三篇数学论文，其中两篇涉及他在代数方程理论方面革命性的工作。1830 年 2 月，他将自己的《论方程可用根式解的条件》一文提交给法国科学院，以角逐该科学院的数学大奖。伽罗瓦的论文，包括柯西在内的许多数学家都曾认为很可能得奖。霉运当头，当时担任科学院常务秘书的傅里叶把论文带回家中，但未及审阅就去世了。结果伽罗瓦的论文再次丢失。1830 年 6 月，法国科学院把这项大奖授予阿贝尔和雅可比，伽罗瓦的论文因为丢失而未能正式参赛，于是，伽罗瓦在这次竞争中莫名其妙地失败了。

命运真是太残酷了。这个不幸的孩子在激怒之中宁愿相信重复发生的事件不是偶然失误而是有意迫害。愤于世道不公的伽罗瓦开始卷入革命的潮流中。1831 年 1 月 3 日，因为激进的革命立场，他被学校正式开除。不能再上学了，他决定用他编写的代数教材自己开班教学生谋生。但不幸的是，他无法吸引住他的学生。于是，他应征加入了国民警卫队的炮兵部队。1831 年 1 月，在数学家泊松的建议下，伽罗瓦把《关于用根式解方程的可解性条件》一文重新提交给法国科学院。与此同时，他继续从事着共和党人的事业。不久他被捕了，但被无罪释放。7 月 11 日，科学院对他的论文做出判决，负责审阅的泊松没能理解伽罗瓦全新的思想，结果他声称伽罗瓦的“论证既不够清晰，又不够详尽，使我们无法判断其严格性”。最后一线希望也破灭了，这一次的受挫基本宣告了伽罗瓦数学生命的终结。他又投身于政治动乱之中。结果他第二次被捕，并被判入狱 6 个月。1832 年 3 月，他被假释。不久，伽罗瓦陷入爱河，这是他的第一次爱情——实则却是继许多悲剧后的又一次悲剧。因为很快他就为此卷入一场无谓的决斗。5 月 29 日，决斗前夕，深信自己将在这场决斗中死去的伽罗瓦写了三封信。最长的第三封信是他的科学遗言，在总共 7 页匆匆写就的这封信中，伽罗瓦整理了自己的数学思想和数学发现。这份理论纲要非常简洁，但内容极其丰富，只是在后来人把它展开的时候，其重要性才渐渐为人所理解。而且提要预见了很久以后的发现，这证明了伽罗瓦深刻的洞察力。信中还有一些后人至今读来仍无比惋惜的文字：“这些

题目并非我已经研究的全部……我没有时间了，我的思想在那个领域未能得到充足的发展——那个领域是巨大的……我不得不承认，在我有兴趣的领域里，我已宣布但尚未证明、从而使人怀疑的定理确实是太多了。”其中最令人难忘，也是最悲伤的话语是“我没有时间了”。

在信的末尾，他要求他的朋友“请公开请求雅可比或高斯就这些定理的重要性（不是就定理的正确与否）发表他们的看法。我相信最终会有人发现，将这一堆东西解释清楚对他们是有益的”。

1832 年 5 月 30 日，星期三的早晨，在决斗中他受了重伤，一个农民发现了他并把他送到医院。他对来陪伴他的弟弟说：“不要哭。在 20 岁就死去，需要我全部的勇气。”次日上午 10 时伽罗瓦停止了呼吸。关于决斗的原因，有各种不同的说法流传。历史学家曾争论这场决斗是一场悲惨的爱情还是政治迫害造成的，但无论是哪一种，一位杰出的数学家在他 20 岁时被杀死了，他研究数学才只有 5 年时间，而其全部工作仅仅写在 61 页小纸上。



后来，他的朋友与弟弟尽职地把他的数学论文抄送给高斯、雅可比及其他一些数学家，但是这些数学家的反应却无案可查。1832 年，他的工作在《百科评论》上发表了，但没有对当时的数学发展产生影响。

直到 1843 年，数学家刘维尔开始认识到伽罗瓦思想的价值。1846 年，他将伽罗

瓦的论文发表在极有影响的《纯粹与应用数学杂志》上。在对伽罗瓦论文的介绍中，刘维尔对为什么这位年轻数学家会被他的长辈们拒绝做了反思：“过分地追求简洁是导致这一缺憾的原因。人们在处理像纯粹代数这样抽象和神秘的事物时，应该首先尽力避免这样做。事实上，当你试图引导读者远离习以为常的思路进入较为困惑的领域时，清晰性是绝对必需的……但是现在一切都改变了，伽罗瓦再也回不来了！我们不要再过分地作无用的批评，让我们把缺憾抛开，找一找有价值的东西……我的热心得到了好报，在填补了一些细小的缺陷后，我看出了伽罗瓦用来证明这个定理的方法是美妙和完全正确的，在那个瞬间，我体验到一种强烈的愉悦。”人们终于承认了“19 世纪数学中一位最悲惨的英雄创造的杰作”。伽罗瓦生前几乎不为人所理解的工作，在其死后十几年终于焕发出灿烂的光辉。

下面，就让我们去简单领略一下这位数学天才迸发出的天才思想吧。

光辉的证明

为理解伽罗瓦美妙的思想，我们先来介绍几个相关的概念。

第一个概念是由阿贝尔已经提出过的数域：一个至少含有两个数的数集 F，如果对数的加减乘除（除数不为零）四种运算是封闭的，即 F 中任意两个数的四则运算还是 F 中的数，则 F 称为数域。另外，若数域 F 中的数都属于数域 K，则 F 称为 K 的子域，K 为 F 的扩域。

按照数域的定义，显然，有理数、实数、复数都构成数域。而且容易知道，任意

数域都包含有理数域；也就是说，有理数域是最小的数域，而实数域、复数域都是有理数域的扩域。除这几个最为常见的数域外，还有无数其他数域。我们这里仅举后面章节中会用到的一类数域为例。

一切形如 $a+b\sqrt{2}$（其中 a、b 为有理数）的实数的全体构成一数域。证明这一点，只需说明任意两个这种类型数的和差积商结果仍具有这种形式。此外，容易明白这一数域包含有理数域，因此它是有理数域的扩域。同样，一切形如 $a+b\sqrt{3}$（其中 a、b 为有理数）的实数的全体也构成一数域。一般地，设 F 是一个数域，$m_1 \in F$，$\sqrt{m_1} \notin F$，则 $F(\sqrt{m_1})=\left\{a+b\sqrt{m_1}: a,\ b \in F\right\}$ 仍是一个数域，而且它是 F 的（二次）扩域。

进而，$F(\sqrt{m_1})$（简记为 F_1）可继续扩张：设 $m_2 \in F_1$，$\sqrt{m_2} \notin F_1$，则 $F(\sqrt{m_1},\sqrt{m_2})=\left\{a+b\sqrt{m_2}: a,\ b \in F_1\right\}$ 仍是一个数域，我们把它简记为 F_2，则它是 F_1 的（二次）扩域。根据域论知识，可知它还是 F 的四次扩域。按照这种方式，我们可逐步扩张：$F \subset F_1 \subset F_2 \subset \cdots F_n$，其中 $F_n=F(\sqrt{m_1},\ \sqrt{m_2},\ \cdots,\ \sqrt{m_n})$，它称为 F 上的 n 级二次扩张塔。下面，我们举一个例子来理解一下这种扩域的形成方式。

考察 $x=\dfrac{\sqrt{5}-1+\sqrt{-2\sqrt{5}-10}}{4}$。令 $F_0=Q$，$5 \in F_0, \sqrt{5} \notin F_0$，则 $F_1=F(\sqrt{5})=\left\{a+b\sqrt{5}: a,\ b \in F_0\right\}$ 是一个数域，且它是 $F_0=Q$ 的二次扩域。然后，取 $-2\sqrt{5}-10 \in F_1$，$\sqrt{-2\sqrt{5}-10} \notin F_1$，则 $F_2=F(\sqrt{5},\ \sqrt{-2\sqrt{5}-10})=\left\{a+b\sqrt{-2\sqrt{5}-10}: a,\ b \in F_1\right\}$ 仍然是一个数域，且它是 $F(\sqrt{5})$ 的二次扩域，是 $F_0=Q$ 的四次扩域。而我们考察的 $x \in F_2$。

有了域与扩域的概念后，我们再来介绍一下数域上的不可约多项式。

系数都是 F 中的数的多项式，称为数域 F 中的多项式；对于数域 F 中的多项式 $p(x)$，如果存在数域 F 中的多项式 $p_1(x)$ 和 $p_2(x)$，使得 $p(x)=p_1(x)\ p_2(x)$，且 $p_1(x)$ 和 $p_2(x)$ 的次数都至少是一次的，则称多项式 $p(x)$ 为数域 F 中的可约多项式；反之，若 $p(x)$ 不能分解成 F 上两个次数较低的多项式乘积，则它称为数域 F 中的不可约多项式，也称既约多项式。

由这一定义可以明白，不能孤立地谈一个多项式是否可约，而应相对于给定的数域而论。比如，$p(x)=x^2-2$，系数都属于有理数域Q，但它在Q中不能分解为两个因式的乘积。因此，它是有理数域Q中的不可约多项式。但是，正如我们已经熟知的，$p(x)=x^2-2$在有理数域的扩域实数域中是可约多项式。事实上，因为$p(x)=x^2-2=(x+\sqrt{2})(x-\sqrt{2})$，这表明它在数域$Q(\sqrt{2})$上即可分解，因此它在有理数域的扩域$Q(\sqrt{2})$中是可约的。

下面再来介绍“群”的概念。群可以理解为一类集合G，集合中元素之间存在一种运算$\times$，这种运算满足下列4条性质。

- （封闭性）对任意a、$b\in G$，都有$a\times b\in G$；
- （结合律）对任意a、b、$c\in G$，都有$(a\times b)\times c=a\times(b\times c)$；
- （单位元）对任意$a\in G$，存在$1\in G$，使$a\times 1=1\times a=a$；
- （逆元）对任意$a\in G$，存在$a^{-1}\in G$，使$a\times a^{-1}=1$。

需要说明的是，上面运算用$\times$表示，只是一种习惯。事实上，群中的运算是任意的，它叫什么名称，或者用什么符号表示都是可以的。同样地，上面的单位元采用的符号1（群中的单位元也常用符号I表示），也只是一种习惯，它并不代表我们所熟知的自然数1。对于不同的群而言，单位元其实是极不相同的。我们举一个简单例子，整数集关于通常意义上的加法运算即构成一个群。显然，封闭性（任意两个整数的和仍是整数）与结合律是满足的。这个群的单位元是整数0（任何整数与0的和不变），而这个群中任意一个整数的逆元是其相反数（任何整数与其相反数的和等于该群的单位元0）。

有了抽象群的概念后，我们再来看一类特殊的群：置换群。有必要补充一句，在数学的发展中，抽象群的概念是在置换群及更多种类群的发现之后才提出的。

为了解置换群，先要说明一下什么是置换。让我们从二次方程谈起。如我们所熟悉的，二次方程 $x^2+px+q=0$ 有两个根 x_1, x_2。由这两个根可以组成许许多多的表达式，比如 x_1+x_2, x_1x_2, x_1-x_2 等。在这些表达式中，交换 x_1, x_2 的次序，即把 x_1 换为 x_2，同时把 x_2 换为 x_1，这就是一个置换。特殊地，把 x_1 换为 x_1，同时把 x_2 换为 x_2，这也是一个置换，称为恒等置换。置换还可以用另一种角度理解。x_1, x_2 的排列有两种：x_1x_2 与 x_2x_1。从一个排列变到另一个排列，这一过程称为置换。显然，在这种情况下，可以得到两个置换，即 x_1x_2 变为 x_2x_1，以及 x_2x_1 变为 x_1x_2。这两个置换可以分别表示为 $\begin{pmatrix} x_1x_2 \\ x_2x_1 \end{pmatrix}$ 与 $\begin{pmatrix} x_1x_2 \\ x_1x_2 \end{pmatrix}$。为简便，置换也可以考虑只用其下标，于是它们可分别表示为 $\begin{pmatrix} 12 \\ 21 \end{pmatrix}$ 与 $\begin{pmatrix} 12 \\ 12 \end{pmatrix}$。更为简便的是，$\begin{pmatrix} 12 \\ 21 \end{pmatrix}$ 可以记为（12），其中括号的数字是环状排列的，意思是 1 变为 2，2 变为 1；而恒等置换 $\begin{pmatrix} 12 \\ 12 \end{pmatrix}$ 则可记为（1）。

再以三次方程三个根 x_1, x_2, x_3 为例。这三个根的不同排列共有 6 个，相应的置换也有 6 个。它们用简单的记法可表示为：（1）（12）（13）（23）（123）（132）。解释一下，（1）表示的是恒等置换。而（12）表示 1 变为 2，2 变为 1，3 变为自身 3，这种自己保持不变的情况，可以在表示中直接省略。于是可以明白，（13）表示的是 1 变为 3，3 变为 1，而 2 变为自身 2。而（123）表示的是 1 变为 2，2 变为 3，3 变为 1。

两个置换之间可以做“乘法”。比如说，（123）·（23）代表的意思是，先作置换（123），再作置换（23）。于是这个乘积可如下求得：先看 1，由（123）知，1 变为 2，又由（23），知 2 变为 3，因此结果是 1 变为 3；同样，2 变为 3，3 变为 2，因此结果是 2 变为 2；而 3 变为 1，1 变为 1，所以 3 变为 1。因此，最后乘积的结果可以记为（13）。类似地，可以计算，（23）·（123）=（12）。通过这例子可以看到，置换的乘法运算不满足交换律。

现在，我们把这 6 个置换看作一个集合，把上面的“乘法”作为一种运算，于是，我们可以得到一个群。其一，集合中任两个置换（元素）乘积的结果仍在这一集

合内，即这个集合关于乘法运算是封闭的。其二，容易验证结合律也是成立的。其三，恒等置换（1）是单位元。其四，任何一个置换都有逆元。比如，（123）的逆元是（132）。由此得到的群就称为置换群。因为群中元素的数目叫群的阶，于是，我们说集合 $\{1, 2, 3\}$ 中 3 个对象的置换全体构成一个阶为 $3!=6$ 的置换群，并可记其为 S_3；类似地，2 个对象 $\{1, 2\}$ 的置换全体构成一个阶为 $2!=2$ 的置换群 S_2。一般地，n 个不同对象的所有置换共有 $n!$ 个，它们的全体构成一个群，这个群通常记为 S_n。

在伽罗瓦的证明中，用到的就是这种置换群。除此之外，还有另一个伽罗瓦用到的有力数学工具，就是子群的思想。如果一个群的某些元素也可以构成一个群，那么这个群称为原来群的子群，原来的群称为母群。自然，任何一个群都至少有两个子群：它本身及由单位元一个元素组成的群，这两个群称为平凡子群。比如，对 S_2 来说，它只有两个平凡子群。对 S_3 来说，它除了有两个平凡子群外，还有一些子群，比如，(1), (1, 2, 3), (1, 3, 2) 这三个置换可组成 S_3 的一个 3 阶子群。

对一个有限阶的群来说，母群与子群的阶具有一种特别的关系，此关系被称为拉格朗日定理，即若设母群 G 的阶为 n，子群 H 的阶为 m，则 m 必整除 n。而所得商数 $\frac{n}{m}$ 叫 H 在 G 中的指数，记为 $[G:H]=\frac{n}{m}$。

有了以上这些铺垫之后，我们可以走进伽罗瓦大脑的迷宫，去欣赏一下他的极有创造性的思想了。

让我们还是从比较简单的例子入手。我们上面已经提到 $p(x)=x^2-2$ 系数都属于有理数域 Q，但它是有理数域 Q 中的不可约多项式，然而它在有理数域的扩域 $Q(\sqrt{2})$ 中是可约的。换一种等价的说法是，方程 $p(x)=x^2-2=0$ 的系数属于有理数域 Q，但方程在 Q 中无解，不过它在 Q 的扩域 $Q(\sqrt{2})$ 上有解。这个简单例子说明，一个方程在其系数所属的域（称为基本域）中无解，却可以在基本域的某个扩域中有解。我们把能将一个方程的所有解都包含在内的基本域的最小扩域称为此方程的根域或分裂域。显然，$x^2-2=0$ 的基本域是 Q，而其根域是 $Q(\sqrt{2})$。再以 $p(x)=x^2+x-1=0$ 为例，它的基

本域是有理数域，在其上方程没有解，但在扩域 $Q(\sqrt{5})$ 上有解。事实上，我们知道这一方程的两个解为 $x=\frac{-1\pm\sqrt{5}}{2}$。对一般的二次方程 $x^2+px+q=0$ 而言，其基本域可记为 $Q(p, q)$。设其在基本域上不可约，那么在这个域中无法得到方程的解。但若把量 $\sqrt{p^2-4q}$ 加入到基本域中，则在得到的扩域 $Q(p, q, \sqrt{p^2-4q})$ 中方程可解，事实上其解可以表示为：$x_{1,2}=\frac{-p\pm\sqrt{p^2-4q}}{2}$。

下面再来考察三次方程 $x^3+mx=n$，在它的基本域 $Q(m, n)$ 上一般无法求解。但我们分别把量 $\sqrt{\frac{n^2}{4}+\frac{m^3}{27}}$ 与 $\sqrt[3]{\pm\frac{n}{2}+\sqrt{\frac{n^2}{4}+\frac{m^3}{27}}}$ 加入到原来的域中，于是在得到的扩域 $Q\left(m, n, \sqrt{\frac{n^2}{4}+\frac{m^3}{27}}, \sqrt[3]{\pm\frac{n}{2}+\sqrt{\frac{n^2}{4}+\frac{m^3}{27}}}\right)$ 中方程可解。事实上，用前面介绍过的卡尔丹公式，我们知道 $x=\sqrt[3]{\frac{n}{2}+\sqrt{\frac{n^2}{4}+\frac{m^3}{27}}}-\sqrt[3]{-\frac{n}{2}+\sqrt{\frac{n^2}{4}+\frac{m^3}{27}}}$。

通过上面的例子，我们看到能用根号求解方程的求根公式特点是，一个代数方程（数字系数或字母系数）有根式解，相当于系数域 F 可以经过有限次添加根式而一扩再扩，直到扩充至根域 K 为止。或者说，相当于在系数域 F 与根域 K 之间存在一个有限的扩域序列：$F=K_0\subset K_1\subset K_2\subset\cdots\subset K_n=K$，其中每一个 K_i 都是由 K_{i-1} 添加 K_{i-1} 中的数的根式而成的扩域。

于是，根式解问题可以转化为域的问题。然而，这种转化中的关键点在于，如何由基本域通过不断扩张最终得到根域。为了解决这一困难，伽罗瓦引入了重要思想，即把域的问题转变成群的问题。我们举简单的例子来说明一下。

考虑方程 $p(x)=x^2+x-1=0$，其系数域为有理数域 Q。由韦达定理可知，$x_1+x_2=-1, x_1x_2=-1$，这是以系数域（此处为 Q）中的元素为系数，由方程的根形成的两个代数关系。容易明白，这两个代数关系在置换（1）、（12）下都保持不变。事实上，可以证明，以系数域中的元素为系数，方程的根所形成的全部代数关系，在这两个置换下都将保持不变。于是，我们说方程 $p(x)=x^2+x-1=0$ 在有理数域中的群由这两个

置换构成。再考虑$p(x)=x^2+x-1=0$在扩域$Q(\sqrt{5})$中的情况。在这种情况下，我们可以考察x_1-x_2。易知，这一代数关系在置换（12）下会发生改变。但显然，它在恒等置换（1）下保持不变。于是，我们说方程$p(x)=x^2+x-1=0$在$Q(\sqrt{5})$中的群由置换（1）构成。

通过这个简单的例子，我们只是想点明几点。其一，对一个方程，我们可以引入伽罗瓦所称的“方程的群”（现在又称“伽罗瓦群”）的概念。所谓“方程的群”，是指一个方程的根形成的置换群中某些置换组成的“子群”。其二，扩大数域以后，方程的伽罗瓦群可以变小。其三，需要指出，保持根的代数关系不变，就意味着在此关系中根的地位是对称的。因此，伽罗瓦群作为方程的“对称外形”，刻画了方程根的对称性。特别地，一个方程的伽罗瓦群是方程假定解的最大置换群。具体来说，n次方程具有n个解。n个解的可能置换的最大数是$n!$，而包含所有置换的群就是群S_n。伽罗瓦证明了，对任何次数n，总能找到一些方程，其伽罗瓦群为整个S_n，比如，存在五次方程的伽罗瓦群是S_5。其四，一个方程在某数域中的群总是可以求出的。可以注意到，上面简单例子中，我们只是给出了结果，并没有认真解释结果的由来。限于篇幅，这里不打算对此做更多介绍。

更重要的一点是，我们可以看到在伽罗瓦域与伽罗瓦群之间存在一种对应关系，这种关系在一般情况下也是成立的。具体而言，设方程在系数域F上对应的伽罗瓦群为G，于是，介于方程的系数域F与根域K之间的所有域恰好与群G的所有子群之间具有一一对应的关系。这被称为伽罗瓦扩域基本定理，整个伽罗瓦理论的核心就是这个基本定理，它是伽罗瓦理论的顶峰。原因在于，域是很复杂的研究对象（它里面包含四种运算和无穷多个数），而置换群（它里面只包含一种运算和有限多个元素）的问题相对而言容易解决。如有人比喻，如果把域的结构问题比作一座高山，从某种角度来看简直是悬崖峭壁，无处攀缘，从置换群的角度来看，山穷水尽之余却是柳暗花明的世界。而伽罗瓦的这一伟大发现表明，群的结果可反映域的结果，从而有可能通过

研究较简单的置换群来解决更复杂的域的问题。

至此，我们看到作为一系列独创性思想的第一步，也是关键的一步，伽罗瓦先把研究代数方程过程中遇到的问题转化成域的问题，然后又把域中遇到的困难转化为群中一系列新问题，由此，全部问题归结为置换群及其子群结构的分析。关于群的研究，同样闪耀着伽罗瓦的天才与非凡的想象力与创造力。

首先，伽罗瓦在子群基础上引入了一种特殊的子群：正规子群。设 G 是群，H 是 G 的子群，任取 $h \in H, g \in G$，若有 $g^{-1}hg \in H$，则称 H 是 G 的正规子群。$g^{-1}hg$ 称为 h 应用 g 的变形，上述定义表明，用任何 $g \in G$ 对 H 中元素变形仍在 H 中，即 H 对变形有“不变”性，因此正规子群也称为不变子群。

一般来说，如果一个群具有正规子群，其中一个将是最大的(即有最高阶，或者说有最多的元素)。进一步，这个最大正规子群可能又有它自身的子正规子群，其中一个又有最高的阶。于是，我们可得到最大正规子群的一个完整系列：$G \supset G_p \supset G_{p-1} \supset G_{p-2} \supset \cdots G_2 \supset G_1$。在得到这个最大正规子群的完整系列后，我们可以进一步做如下定义：令 $K_1 = \frac{n}{n_p}$, $K_2 = \frac{n_p}{n_{p-1}}$, $K_3 = \frac{n_{p-1}}{n_{p-2}}$, $\cdots$, $K_p = \frac{n_2}{n_1}$（其中，n 是群 G 中含有的元素数，n_p 是群 G_p 中含有的元素数，n_{p-1} 是群 G_{p-1} 中含有的元素数……n_1 是群 G_1 中含有的元素数，易知 G_1 为单位元群，因此 $n_1 = 1$)，由此得到的指数序列 $K_1, K_2, \cdots, K_p$ 称为群 G 的组合因数。为方便理解，我们举例子说明一下。

先以置换群 S_2 为例。S_2 只有一个子群即单位元群(1)，显然这个子群也是正规子群。于是，我们得到的最大正规子群系列很简单：(1)和(12)；(1)。于是，置换群 S_2 的组合因数为：$\frac{2}{1} = 2$。

再看一下置换群 S_3，这个群包含6个元素：(1)、(12)、(13)、(23)、(123)、(132)。可以验证，含有(1)、(123)、(132)3个元素的群H是其正规子群，且是最大正规子群。进一步，H 的最大正规子群是单位元群(1)。于是，我们得到 S_3 的最大正

规子群系列是 S_3 、H、单位元群。置换群 S_3 的组合因数为 $\frac{6}{3}=2,\ \frac{3}{1}=3$ 。

在群 G 组合因数概念基础上，伽罗瓦引入了可解群概念：若群 G 中的组合因数 $K_1, K_2, \cdots, K_p$ 全都是质数，则 G 称为一个可解群。“可解”之名来源于代数方程的可解性。事实上，经过千曲百折之后，我们已经走到了迷宫的出口。这就是伽罗瓦光辉的定理：一个方程有根式解的充分且必要条件是，这个方程的群是可解群。具体而言，如果一个方程在一个含有其系数的数域中，它的群是一个可解群，则此方程有根式解，而且仅在这一条件下方程有根式解。

下面我们应用伽罗瓦这一“美妙的定理”，来判定代数方程的可解性。

对于一般的二次方程 $ax^2+bx+c=0$ ，它有两个根 x_1, x_2 。此方程在一个含有其系数的数域中，它的群 S_2 中的元素只能是（1）和（12）。如上所述，置换群 S_2 的组合因数为 2。2 是质数，这表明群 S_2 是可解群。因此，由伽罗瓦定理，二次方程都有根式解。

对于一般的三次方程 $ax^3+bx^2+cx+d=0$ ，它有 3 个根 x_1, x_2, x_3 。此方程在一个含有其系数的数域中，它的群 S_3 中的元素包含 6 个元素：（1）、（12）、（13）、（23）、（123）、（132）。如前所述，置换群 S_3 的组合因数为 2、3。因为 2 与 3 都是质数，这表明群 S_3 是可解群。因此，根据伽罗瓦定理，三次方程都有根式解。

对于一般的四次方程 $ax^4+bx^3+cx^2+dx+e=0$ ，它有 4 个根 x_1, x_2, x_3, x_4 。此方程在一个含有其系数的数域中，它的群 S_4 中包含 $4!=24$ 个元素。可以验证，这个群的最大正规子群 G_4 为 12 阶，而 G_4 的最大正规子群 G_3 为 4 阶， G_3 的最大正规子群 G_2 为 2 阶， G_2 的最大正规子群为 1 阶。因此，其组合因数是 2、3、2、2。它们都是质数，这表明群 S_4 是可解群。于是，根据伽罗瓦定理，四次方程都有根式解。

这样，我们应用伽罗瓦定理证明了人们已经熟悉的结论：不超过 4 次的代数方程都有根式解。下面我们继续考察一下 5 次及高于 5 次的代数方程。

为此，伽罗瓦首先证明了一个结论：当 $n>4$ 时， n 次代数方程在一个含有其系数

的数域中，它的群除了单位元群 1 以外，只有一个不变子群 A_n，且其元素个数为 $\frac{n!}{2}$。于是，对一般的 n 次代数方程而言，其根有 n 个。所以在一个含有其系数的数域中，它的群 S_n 含有 $n!$ 个元素，由伽罗瓦证明的结论，此群的最大不变子群为 A_n，而 A_n 的最大不变子群是（1），所以组合因数为 $n!\div\frac{n!}{2}=2, \frac{n!}{2}\div 1=\frac{n!}{2}$。2 是质数，但容易知道 $\frac{n!}{2}$ 在 $n>4$ 时一定不是质数，这表明，S_n 在 $n>4$ 时不是可解群。于是，根据伽罗瓦定理，当 $n>4$ 时，一般的代数方程没有根式解。

多么简练、完美、奇妙的解决！异常简练，达到了不能再简练的地步；异常完美和奇妙，达到了无法更完美、更奇妙的程度。伽罗瓦令人拍案惊奇、叹为观止的解答，使代数方程的根式求解，这个在数学史上让人深感兴趣的问题终于得到圆满而彻底的解决。

结　　语

在本章，我们介绍了人类在代数方程根式求解方面的不懈努力与最终成功。让我们再简单回顾一下人类为此走过的历程。

早在河谷文明时期，人类（特别是古巴比伦人）已掌握了一次与二次方程的求解法则。到中世纪，阿拉伯数学家特别是花拉子密又将二次方程的理论系统化。文艺复兴时期及以后一段时间，欧洲数学家又在两方面获得了突破：一方面，成功解决了三次和四次代数方程的求根问题；另一方面，创建并逐步完善了符号代数。由此，次数不超过 4 的代数方程的解法都可以写成一个代数表达式，这个表达式可以由方程的系数

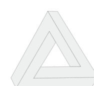
通过有限次的有理运算（即加减乘除运算）或开方而得到。因此，我们说四次的一般代数方程都有根式解。

非常自然地，人们希望把这一结论推广到一般的五次及更高次方程中去，即获得五次或更高次代数方程的根式解。然而，两个多世纪的尝试都以失败告终。直到 1770 年，事情开始出现了转折。法国数学家拉格朗日在研究中发现适用于不超过四次代数方程的方法并不适用于解更高次方程，于是他猜测高次方程不能根式求解。另外在其探索中，他还提出置换的思想，并意识到置换理论是“整个问题的真正哲学”，但他未能继续前进。1799 年，意大利数学家鲁菲尼提出并证明了五次方程没有根式解，但其复杂又有缺陷的证明未得到数学界的重视。1824 年，年仅 22 岁的数学家阿贝尔迈出了决定性的一步，他严格证明了五次及高于五次的一般方程没有根式解。在这一非凡结果之后，剩下的问题是：什么样的特殊方程能够用根式来求解？在这方面，高斯研究并证明了分圆方程有根式解。在高斯研究基础上，阿贝尔也考虑了一些特殊的能用根式求解的方程，并在其工作中引进了“域”的思想。阿贝尔因早逝而留下的未竟之业留给了另一位与其命运相似的年轻数学家。在 1829~1831 年完成的几篇论文中，伽罗瓦的天才一笔抒写了判别方程根式可解的充分必要条件，从而为方程根式可解性这一经历了三百年的难题画上了圆满的句号。

通过本章的介绍，我们看到，在一代代数学家所关注的这一基本问题的漫长求索过程中，古典代数学得到极大丰富与发展。更为重要的是，跑最后一棒的伽罗瓦为解决问题而创造性地引入了全新的“群”概念与“群论”思想。这开辟了代数学的新源泉，引发了代数学的革命性变化。在伽罗华之前，解方程始终占据着代数学舞台的中心，代数学曾只被认为是解方程的学问，而在他之后，代数学被引入一个新的轨道，代数学的面貌焕然一新。代数学不再限于研究方程，而是更多地把注意目标转向各种代数结构。由此，代数学跨过解方程的阶段，进入研究新的数学对象（群、环、域、格等）的阶段，逐渐形成数学的重要分支“近世代数”，又称“抽象代数”。追根溯源，

群的引入正是在方程根式解这一古典问题的深入研究中获得的。因此，我们可以说，“数学抽象的顶尖艺术”——群论——是方程根式解问题带给数学的最伟大收获，一颗最大的金蛋。

我们还可以体会到一点，数学的抽象不是来自数学家的想象，而是来自成功攻克从经典数学继承下来的问题的需要。抽象的结构数学不是无源之水、无本之木。它的产生不但大大扩大了数学对象的领域，而且与经典数学结合在一起，对经典数学问题的解决有着极大促进作用。在下一章，我们就会进一步体会到群论这一“美丽而且深刻的智力创造物”作为一种“处理范围极广数学问题的无法估价工具”的伟力。在第三章，我们还会看到群概念在澄清和统一不同数学分支方面的作用。



第二章

几何三大问题

第一节
几何三大问题的由来

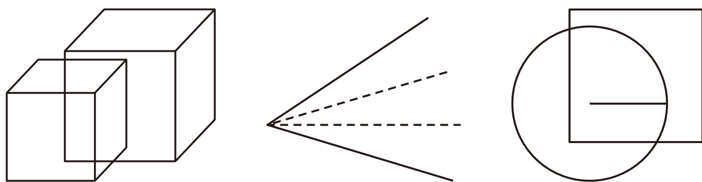
公元前5世纪，古希腊雅典出现了一个包括各方面学者的诡辩学派（也称智者学派）。他们第一次提出并研究了三个作图问题：倍立方（体）问题、三等分角问题以及化圆为方问题。这就是数学上著名的具有历久不衰魅力的几何三大问题。这三大问题的共同点是寻求一种仅用直尺和圆规的解法。在这一限制条件下，三大问题虽一直吸引着人们的注意，但历经2000多年却未能真正解决。直到19世纪，人们才最终证明了找不到解法是因为解法根本不存在，也就是说这三大问题是尺规不能问题。本章中我们要讲述的正是三大几何问题从起源到被证明不可解的曲折而漫长的历史。

在本节中，我们将首先介绍几何三大问题的由来及尺规作图的来历。事实上，尺规作图的要求正是三大问题难解的症结。

几何三大问题的由来

几何三大问题的意思非常容易理解。倍立方问题是说，求作一立方体，

使其体积等于一已知立方体体积的二倍。三等分角问题是说，三等分任意角。化圆为方问题是说，求作一正方形，使其面积等于一个已知圆的面积。



对这三个吸引了人们2000多年的著名问题，我们先来了解一下它们是如何起源的。

三者中，倍立方问题的起因说法多而且有趣。

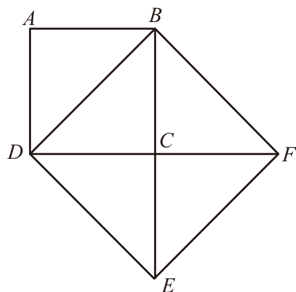
一种传说与克里特岛的米诺斯王有关。米诺斯王在古希腊神话中是宙斯之子，后成为克里特岛著名的君王。他死后成为地府的三个判官之一，决定阴魂的未来命运，惩罚有罪的灵魂。后来一位古希腊诗人在一个故事中叙述说："米诺斯发现世人为他的儿子格劳库斯建造的立方体的墓边长只有100尺，认为这'太小了，你们要建造的墓可是高贵王族的栖息之所啊。墓要加大一倍'"。这就是所谓的倍立方问题。这位诗人继续传达米诺斯的吩咐："不改变墓的形状，尽快将墓的各边长加倍——'只要将每边增大一倍，就可以实现我的要求'。"这当然是不对的，因为每边加倍，体积将会变为原来的8倍。

还有另一种关于倍立方问题由来的传说。公元前430年，希腊雅典发生了瘟疫，相当多的希腊人失去了生命。为了消除灾难，得洛斯岛的公民代表被派往特尔斐的阿波罗神殿求助，太阳神阿波罗降下神谕，说必须把他的立方祭坛增大一倍且形状仍为立方体，疫病才不会流行。一位自作聪明的设计师将祭坛的每边增大一倍，做了一个新的祭坛，放在太阳神庙里，结果太阳神大怒，疫势反而更加猖獗。因为新祭坛的体积是原来的8倍，而不是2倍。那么应当怎样做才符合要求呢？这就是著名的倍立方问题，也称为得洛斯问题。又据说，德利安就这个问题请教过古希腊哲人柏拉图，所

以这一问题也称德利安问题。没能解决这一问题的柏拉图对神谕的真意给出一个有个性的回答。他说，神这样答复，他并不是想得到一个体积加倍的祭坛，而是希望通过派给希腊人这项工作，来责罚希腊人对于数学的忽视和对几何学的轻视，由此重新唤起希腊人对纯几何研究的兴趣。这里把柏拉图拉进来显然是错误的，因为这场瘟疫发生的时候，柏拉图才刚出生。

在一部阿拉伯著作中，希腊稗史又被转述成如下的说法："柏拉图在世的时候，以色列儿童中爆发一场瘟疫。于是神向一位先知传来的声音说'立方祭坛加一倍，瘟疫鬼头即粉碎'。于是众人另做了一个和原来同样的祭坛，把它和旧祭坛并放在一起。然而疫情蔓延日甚一日。这时神又向先知说'两个祭坛排成队，岂乃立方体加倍？体积虽倍形长方，亵渎我神尔等罪！'于是众人向柏拉图请教。柏拉图告以'几何学为众科学中至为崇美卓绝之学科，尔等竟无视之，神灵以故乃施惩罚。'"

这些对于问题传播起了推波助澜作用的传说自然是无稽之谈。实际上，古希腊人想到这一问题是非常自然的。古希腊人熟知，以正方形的对角线为一边作一个正方形，其面积是原正方形的两倍（如下图所示），理所当然会提出相应的立体问题，即倍立方问题。



另外两个著名的作图问题也是希腊人在解出了一些作图题之后的自然引申：因任意角可以二等分，于是就想搞三等分；解决了一些具有一定形状的图形使之与给定图形等（面）积的作图题之后，想到圆和正方形是最简单的几何图形，自然地提出化圆为

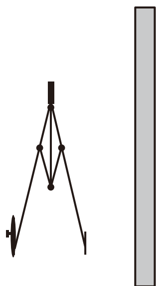
方问题。

介绍过三大问题的由来后，为避免对这三大问题产生误解，我们有必要强调对这三大几何作图问题的要求。

一个要求是：从理论上精确地解决这个问题。古希腊人关心的只是精确的理论解，对问题的近似解不感兴趣，比如说，从实践角度而言，要做出一个近似的两倍体积的祭坛并非难事。另一个要求是：作图只允许用尺规。我们后面会看到，这一限制条件非常关键。事实上，整个问题难解的症结都集中在这一要求上。对此，我们有必要先了解一下为什么古希腊人会对作图提出这一特殊要求。

尺规作图的规矩与来历

学过平面几何的人都清楚，在几何作图中作图工具限用直尺（无刻度）和圆规两种，这一习惯来自古希腊。



为什么古希腊作图只限用尺规呢？原因之一是，这与古希腊人几何研究的对象有

关。古希腊人研究的基本图形是直线和圆，而直尺和圆规是其具体化。有了直尺和圆规这两种作图工具，直线和圆都可作出，自然无须再增加别的工具。另一原因源于古希腊几何的基本精神：从极少的基本假定（定义、公理）出发，推导出尽可能多的命题。对于作图工具，自然也限制到不能再少的程度。此外，古希腊人对尺规作图的严格要求可能受到了柏拉图思想的影响。作为哲学家的柏拉图特别重视数学，他强调几何在训练智力方面的特殊作用。为了达到训练逻辑思维的目的，柏拉图主张作图工具要有所限制，反对用尺规之外的其他机械工具作图。

根据现有的资料，在历史上最先明确提出尺规限制的是伊诺皮迪斯（约公元前465年）。在那以前，许多作图题都是不限工具的。以后经过柏拉图（约公元前427—前347）大力提倡，尺规的限制渐渐成为一种公约，最后总结在《几何原本》一书中，并由公设的形式规定下来。欧几里得提出的公设有5条，与作图有关的有三条。

- 公设1. 由任意一点到另外任意一点可以画直线。
- 公设2. 一条有限直线可以继续延长。
- 公设3. 以任意点为圆心及任意的距离[1]可以画圆。

显然，根据这几条公设，作图工具只能用尺规。而公设1与公设3在明确了只能允许使用尺规两种工具之后，进一步指出了尺规具有的功能。事实上，这两条公设指明了利用直尺和圆规可以并且只能完成的被认可的简单作图，我们分别称这两条约定为作图公法1°与作图公法2°。顺便指出，这两条公法实际上只能用理想的圆规和直尺才能实现，比如，直尺要足够长，圆规的跨度要放得很大又要能收得很小。因此，这是理想化了的作图规则。

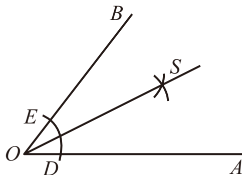
除这两条作图公法外，还有一条自然的约定可作为作图公法3°：两条已知直线，或一条已知直线和一个已知圆，或两个已知圆，如其相交，可确定其交点。此外，作

① 原文中无“半径”二字出现，此处“距离”即圆的半径。

图中还有一个附加规定：在已知直线上或直线外，已知圆周上或圆内或圆外，均可任意取点，但所取的点不得附加其他任何特殊性质。

上面 3 条约定的作图公法，指明了尺规作图的可能范围。于是利用直尺和圆规来完成一个作图题，就是指由上述作图公法所确定的简单作图的有限次组合。下面我们以熟知的二等分角为例来领会一下。

已知$\angle AOB$，求作射线 OS，使$\angle AOS=\angle SOB$。



做法

1）以点 O 为圆心、任意长为半径作弧 DE（根据作图公法 2°，即公设 3），交 OA 于点 D（根据作图公法 3°），交 OB 于点 E（根据作图公法 3°）。

2）分别以点 D 和 E 为圆心、以大于$\frac{1}{2}DE$的长为半径作弧（根据作图公法 2°），这两条弧相交于点 S（根据作图公法 3°）。

3）作射线 OS（根据作图公法 1°，即公设 1）。

则 OS 就是所求作的射线。

可以看到，这一例子的作图过程实质上是作图公法 2°、作图公法 3°、作图公法 3°、作图公法 2°、作图公法 3°、作图公法 1° 的组合。不仅如此，任何一个利用直尺和圆规来完成的作图题，都能转化为作图公法 1°~3° 的有限次组合。凡是能有限次利用 3 条作图公法完成的作图就称为尺规作图可能问题。反之，凡不能有限次利用 3 条作图公法完成的作图就称为尺规作图不能问题。在作图公法的限制下，用尺规确实可以作出许多图形，甚至可以作出相当复杂的图形，但是是否如古希腊人所相信的那

样“不论多么复杂的图形，都能依靠足够的耐心和聪明的才智，仅用尺规把图作出来”呢？并非如此。事实上，有些图形确实无法只用尺规完成，本章介绍的三大作图问题就是最典型的尺规作图不能问题。在长达 2000 多年的时间里，人们未能意识到这一点，为此付出了无数的心血，都以失败告终。但这个探索的过程也孵出了许多数学金蛋。在下一节中，我们就要去了解一下人类这一悲壮却并非无所获的历程。



第二节 几何三大问题的历史解答

几何三大作图问题自提出后，就吸引了许多人投身其中，但由于只能使用直尺和圆规，这使问题变得难以解决并因而更富魅力。虽然许多研究者对三大问题给出了解答，所有这些解答都无法严格遵守尺规作图的限制，但这些研究却引出了大量新发现，它们对整个几何学的发展产生了巨大影响。本节中，我们将分别介绍对几何三大问题的历史解答。

倍立方问题的历史解答

倍立方问题的表述是：已知一个立方体，求作一个新立方体，使其体积是已知立方体的两倍。设已知立方体的棱长为 a，所求倍立方体的棱长为 x，显然满足 $x^3=2a^3$。

在倍立方问题提出后不久，希波克拉底（约公元前460—前370）迈出了解决问题的第一步。作为一项重要的进展，希波克拉底的发现是把这一问题做了简化。他指出，设原立方体的棱长为 a，那么只要能作出已知线

段 a 与 $2a$ 的“双比例中项”x、y，那么 x 即为所要求的新立方体的棱长。所谓 a 与 $2a$ 的“双比例中项”x、y，即指 x、y 满足比例式子：$\frac{a}{x}=\frac{x}{y}=\frac{y}{2a}$。容易看出这与原问题确实是等价的：由 $\left(\frac{a}{x}\right)^3=\frac{a}{x}\cdot\frac{x}{y}\cdot\frac{y}{2a}=\frac{1}{2}$，可得 $x^3=2a^3$。

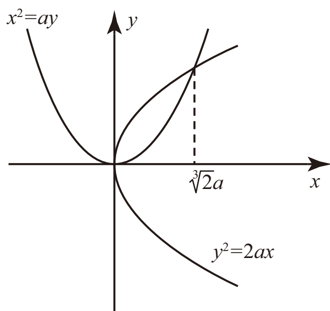
希波克拉底是如何想到把倍立方问题转化为双比例中项的呢？据推测，他是受到了倍正方形问题的启发。我们知道倍正方形问题相当于作出满足 $x^2=2a^2$ 的边长 x，而为得到长度 x，可以做 a 与 $2a$ 的比例中项，即使得 $\frac{a}{x}=\frac{x}{2a}$。因此，很可能这种处理倍正方形问题的方法促使希波克拉底把倍立方问题归结为找出 a 与 $2a$ 之间的“双比例中项”。

希波克拉底没有给出解决倍立方问题的具体方法，只是指出了解决问题的可行途径。经过他的“简化”后，倍立方问题转化成另一个等价的问题，从而使原来的问题变得清晰明了了。在希波克拉底之后，为了解决倍立方问题，人们的探索转向于致力寻找两条已知线段 a 与 $2a$ 的两个中项，即作出满足 $\frac{a}{x}=\frac{x}{y}=\frac{y}{2a}$ 的线段 x、y。在这一思路的启发下，后人确实找到了几种非尺规作图的解法。

门奈赫莫斯解法

公元前 4 世纪的门奈赫莫斯（约公元前 380—前 320）给出了两种做法。一种做法是用现在的解析几何语言表示，即考虑抛物线 $x^2=ay$ 与 $y^2=2ax$ 在第一象限内的交点。

通过解方程组 $\begin{cases}x^2=ay\\y^2=2ax\end{cases}$，得交点的横坐标 $x=\sqrt[3]{2}a$。



另一种做法是，考虑抛物线$x^2=ay$与双曲线$xy=2a^2$的交点，易知其横坐标x、纵坐标y恰好满足上述比例关系。事实上，我们易知交点的横坐标$x=\sqrt[3]{2}a$。

实际上，门奈赫莫斯的巧妙解法正是来自对$\frac{a}{x}=\frac{x}{y}=\frac{y}{2a}$所做的巧妙处理。把这一连比例式拆成两个等式$\frac{a}{x}=\frac{x}{y}$与$\frac{x}{y}=\frac{y}{2a}$，由此得到的两个方程联立求交点，即为第一种解法。同样，如果把这一连比例式拆成两个等式$\frac{a}{x}=\frac{x}{y}$与$\frac{a}{x}=\frac{y}{2a}$，即可得第二种解法。

门奈赫莫斯的解答在现在看来非常自然，但在当时却具有开创意义。一方面，他采用了曲线相交的“作图”方法，这种方法被许多后来者继承。更重要的方面在于，正是通过他的解答，数学中第一次引入了抛物线、双曲线这两种新的曲线。而在此之前，这两种曲线还不为人所知。

后来，门奈赫莫斯又发现，这两条曲线可以作为平面与圆锥面的截线得到。具体而言，用垂直于圆锥母线的平面去截圆锥面，当圆锥顶角是直角（这种圆锥当时称为直角圆锥）时，所得截线为抛物线；顶角为钝角（这种圆锥当时称为钝角圆锥）时，所得截线为双曲线的一支；顶角为锐角（这种圆锥当时称为锐角圆锥）时，截线为椭圆。这三种曲线后来统称为圆锥曲线。因此，正是在解决倍立方问题的过程中，门奈赫莫斯发现了圆锥曲线。

圆锥曲线的发现引起许多数学家的兴趣。欧几里得、阿基米德等都对其做了深

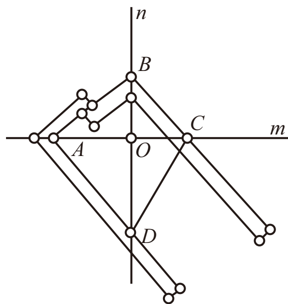
入细致的研究。特别是古希腊伟大的几何学家阿波罗尼奥斯（约公元前 262—前 190）完成了传世巨著《圆锥曲线论》，把圆锥曲线的研究推进到登峰造极的地步。再后来，人们认识到，圆锥曲线不仅在数学中极为重要，而且可应用于物理、天文学，这大大抬高了圆锥曲线的价值。但这一切，追根溯源，都要归于门奈赫莫斯关于倍立方问题的解答。以后人的眼光看，在解决倍立方问题中收获最大的数学金蛋正是圆锥曲线的发现。

当然，抛物线、双曲线这类圆锥曲线是无法用尺规作出的，所以门奈赫莫斯的两种解法给出了倍立方问题的解答，但不合尺规作图的要求。

下面我们再看看其他研究者在这一著名问题上取得的进展。

柏拉图做法

如图，先作两条互相垂直的直线m、n，它们相交于点O，取$OC=a$、$OD=2a$。



然后取两个直角的曲尺（木工常用的工具），使一曲尺通过 C 点，且直角顶点在线 n 上；使另一曲尺通过 D 点，且直角顶点在 m 上。

在上述条件下，移动调整两曲尺的位置，直到两曲尺的另一直角边互相密合为止。这样便在直线 m 与 n 上分别确定了点 A 与点 B。

因为 $OB^2 = OA \cdot OC$，$OA^2 = OB \cdot OD$，所以有 $OC : OB = OB : OA = OA : OD$。

于是 OA、OB 即为所求。

作图过程中使用的曲尺也可用直角板代替。这一借助于机械的解法，通常被称为柏拉图做法。但一般认为这一做法并非由柏拉图给出，因为柏拉图曾对使用尺规之外的机械表示不满，认为这样的结果是一无可取地失去最好的几何学，破坏人们的创造，阻止人们认识永恒的精神世界。

埃拉托塞尼方法

埃拉托塞尼（约公元前 276—前 195），是希腊数学家、地理学家，他兴趣广泛，博学多才。他是著名数学家阿基米德的挚友，约公元前 235 年担任亚历山大图书馆馆长，晚年时由于双目失明而自杀。在数学上他以“素数筛子”的发现者而闻名，在地理学上其脍炙人口的业绩是用简单的方法对地球的周长做了令人不可思议的精确测算。有意思的是，埃拉托塞尼被同时代人戏称为“贝塔”（意即第二），因为他在很多方面的工作都没有达到最高水平。但在现在很多学科领域，人们都公认他是一

位伟大的学者。

在几何方面，埃拉托塞尼叙述了倍立方问题的起源（我们前面介绍过的两种传说就是源自埃拉托塞尼的说法）及求解历史，并对解决这一问题做出了自己的贡献。在下面一首被刻在托勒密三世圣殿大理石板上的短诗中，他夸耀了自己对倍立方问题的解法。

“好朋友，如果你想解决倍立方问题，而不改变立方体的形状。

那么你现在能够如愿以偿。

这样，你就可以加倍一个羊栏、地窖，或者深井的容量。

你只需要在两个直尺之间，捕获顶端共线的两个中项。

你不要去为阿尔希塔斯的圆柱面的活计困惑繁忙，

不要去为门奈赫莫斯的三个圆锥面的截线劳伤，

也不要跟着欧多克索斯去画那些神鬼生畏的曲线形状。

我的方法简单漂亮，你只需滑动这三角形小片数张，

由小到大，你很容易就能找到千千万万个比例中项。

愉快优雅地，你，托勒密！你这青春活力永驻的慈父，

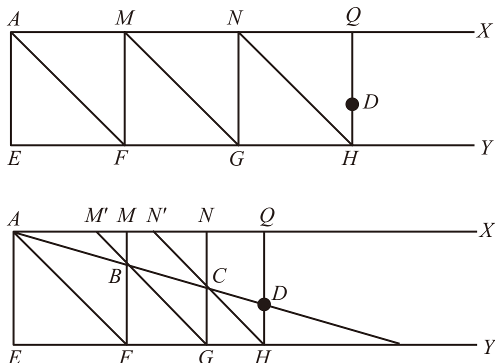
你做的一切，缪斯众女神和那历代列王都要惊讶羡慕。

也许，在将来，啊，宙斯，这上天之主，

也将从你的手中接过这一独特的王杖圣物。

到那时啊，让所有见到这一贡献的人都欢呼：

这正是库热那的埃拉托塞尼的天才赠赋！”



埃拉托塞尼解决倍立方问题的方法借助了一种器械。结合上面的图，我们解释一下他的方法。

如第一图所示，在两条平行线 AX、EY 之间有三个等腰直角三角形 AFM、MGN、NHQ，第一个保持不动，后两个可在平行线上滑动。在 QH 上取一点 D，然后如第二图所示，向左滑动 $\triangle MGN$ 至 $M'GN$ 位置，以使 AD 与 MF 的交点 B 落在其斜边上。同样，向左滑动 $\triangle NHQ$ 至 $N'HQ$ 位置，以使 AD 与 NG 的交点 C 落在其斜边上。

根据 $\triangle AFB \backsim \triangle BGC$，可得 $CG:BF=BF:AE$（相似三角形对应高的比等于对应边的比）；同理，根据 $\triangle BGC \backsim \triangle CHD$，可得 $DH:CG=CG:BF$。

设线段 $AE=2a$，$CG=x$，$BF=y$，并让 $DH=a$（只需取 D 为 QH 的中点），则有 $a:x=x:y=y:2a$。通过上述操作，我们由已知线段 a、$2a$ 得到了其“双比例中项”x、y。根据希波克拉底的结论，倍立方问题得以解决。

为找到埃拉托塞尼所说的千千万万个比例中项，只需在此两平行线间加千千万万个滑动的三角形，依据上述同样的道理即可得到。也就是说，用他的方法，不仅能得到两线段间的两个比例中项，而且能得到所需要的许多比例中项。

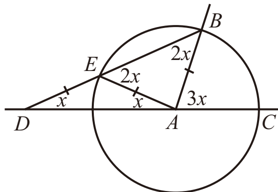
三等分角的历史解答

三等分任意角这一问题何时出现，历史上没有记载。不过，正如我们已提到的，这一问题的提出是非常自然的事情。古希腊人约在公元前 500 年就已经完成了二等分任意角。既然能够容易地二等分任意角，人们自然会考虑把问题引申：能否三等分任意角呢？然而，让人深感诧异的是，由二等分到三等分，仅仅这么一点小小的变化，一个平淡无奇的几何作图题，就变成了一座高深莫测的数学迷宫。下面我们介绍几种走出这一迷宫的方法。

阿基米德方法

古希腊伟大数学家阿基米德（约公元前 287—前 212）在其《引理集》一书中证明了一个命题，后人发现，这一命题事实上可以给出一种三等分角的可能途径。为了叙述方便，我们对阿基米德的原命题稍加修改。

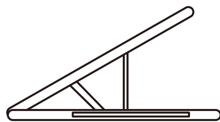
如图，以已知角的顶点 A 为圆心作圆，分别交已知角两边于 B、C。然后过 B 点引直线与圆交于 E、与 CA 延长线交于 D，并且使 DE 的长恰为圆的半径。



可以非常容易地证明，$\angle BDC=\frac{1}{3}\angle BAC$。于是要三等分$\angle BAC$，只需过$A$点引$DB$的平行线即可。

方法非常巧妙而简洁。然而难点在于，如何作出符合条件的线段BED。当然，办法是有的。比如说，在直尺上标出两点，使其长度为圆的半径，然后绕点B转动直尺的位置，使直尺上这两点分别落在圆和CA的延长线上。但这相当于使直尺具有了刻度的功能，与尺规作图中的直尺无任何刻度不符。

事实上，根据阿基米德这一简单明了的原理，可以制作出三等分角的仪器，下图所示即为其中一种。

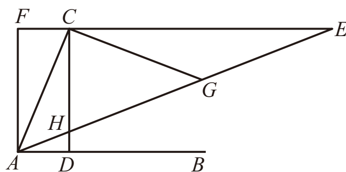


帕普斯方法

希腊亚历山大的帕普斯（约 290—约 350）是罗马帝国晚期的数学家，曾把希腊自古以来名家的著作汇编为《数学汇编》，共 8 卷，其中也包括了他的创造。比如在第 4

卷中，他对三等分任意角问题就给出了自己的解答。为了理解他的方法，我们先对尺规三等分角问题做以下分析。

如图所示，设 $\angle CAB$ 为已知的任意角，而 $\angle HAB$ 是它的一个三等分角，即 $\angle HAB=\frac{1}{3}\angle CAB$。为了作出 $\angle HAB$，我们考虑一个长方形 $AFCD$，FC 的延长线与 AH 延长线相交于一点，设交点为 E。显然，为了作出 $\angle HAB$，只需定出 E 点。作 $\mathrm{Rt}\triangle CEH$ 斜边上的中线 CG，容易证明，$EH=2CA$。



于是，我们得到三等分角的一种方法：设 $\angle CAB$ 为已知的任意角，作一个长方形 $AFCD$，过 A 作一条线，使这条线与 CD、FC 都相交，而且使由这两个交点确定的线段（即图中的 EH）长度恰好为已知线段 CA 的二倍。根据上面的分析，易知 $\angle HAB$ 是 $\angle CAB$ 的一个三等分角。

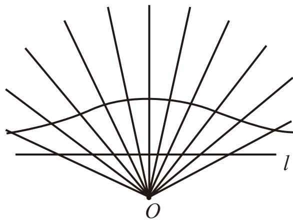
问题在于，如何找到满足条件的 EH，或者说如何确定出点 E 呢？帕普斯借助双曲线实现了这一点。如果借助于有刻度的直尺，也可以做到。具体做法是，在直尺上标出一段长为 $2CA$ 的线段 EH，然后调整直尺的位置，使它过点 A，且 H 落在 CD 上，E 落在 FC 延长线上，即可完成三等分角。但无论是引入双曲线还是使用有刻度的直尺，都明显违反了尺规作图的规定。然而若只用尺规，却无法作出满足条件的 EH，也就无法用帕普斯方法三等分角。

上面两种方法，是古老的，也是最为人熟知的两种三等分角作图方法。

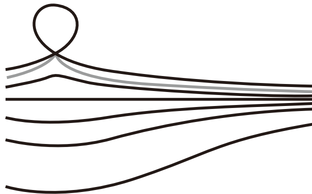
尼科米迪斯的蚌线法

古希腊数学家尼科米迪斯（约公元前 280—前 210）使用蚌线法解决了三等分角问题。为了了解该方法，我们首先介绍一下由尼科米迪斯发现的称为“蚌线”的曲线。

如下图所示，取定点 O 与定直线 l 。通过点 O 作与直线 l 相交的射线，然后在每条这样的射线上以定直线 l 为界向外截取固定的长度，由此确定的动点轨迹即形成蚌线的一支。如果以定直线 l 为界向里截取固定的长度，则可得到蚌线的另一支。其中，定点 O 称为蚌线的极点，定直线 l 称为蚌线的准线。

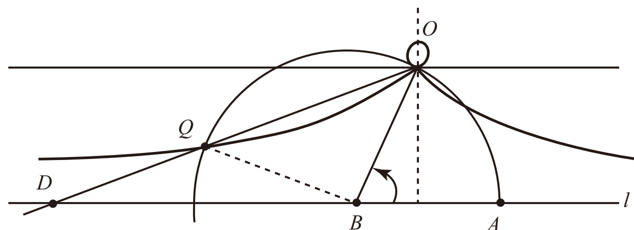


设蚌线定义中截取的固定长度为 a（蚌线定义中的这一固定长度称为蚌线的参数），设定点 O 到定直线 l 的距离为 c ，那么根据 a 与 c 的关系，蚌线可分为三种类型。若 $c > a$ ，蚌线不过 O 点；若 $c = a$ ，蚌线在点 O 处有个尖儿；若 $c < a$ ，蚌线穿过点 O 绕个扣儿又穿回去。如下图所示，每种类型都有两支。



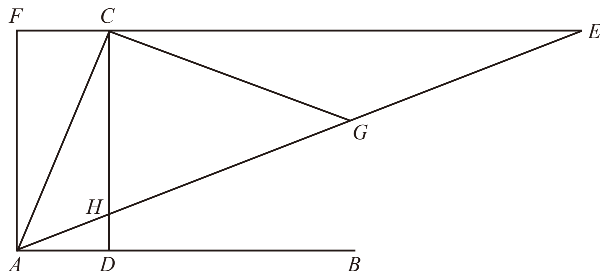
下面就来看看如何利用这种曲线三等分角。如下图所示，要三等分 $\angle ABO$，可以把 O 作为定点，直线 AB 作为准线，BO 之长为蚌线的参数作出蚌线。然后以 B 为圆心作圆，设此圆与蚌线的一支在 $\angle ABO$ 外交于 Q，OQ 与准线交于点 D。

因为 Q 在蚌线上，根据蚌线的定义有 $DQ = BO$，又 Q 在圆上有 $BO = BQ$，因此有 $DQ = BO = BQ$。易知 $\angle BDQ = \frac{1}{3}\angle ABO$。



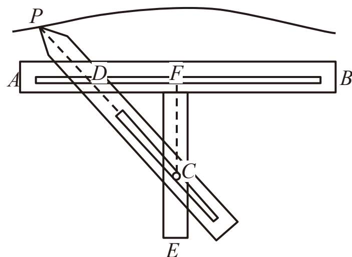
不难明白，这种用蚌线三等分角的办法实质上依据了阿基米德三等分角原理，它把阿基米德方法中无法用尺规确定的点，通过引入蚌线确定下来了。

利用蚌线还可以给出三等分角的另一种办法。如下图所示，设 $\angle BAC$ 为需要三等分的任意角。作矩形 $ADCF$。以 A 为定点、CD 为准线、$2AC$ 为蚌线的参数作出的蚌线与 FC 延长线交于 E，连接 AE。于是按蚌线的定义，$EH=2AC$，作 Rt△CEH 斜边上的中线 CG，易知 $\angle BAE = \frac{1}{3}\angle BAC$。



不难明白，这种用蚌线三等分角的办法依据了帕普斯方法，它把帕普斯方法中无法用尺规确定的点，通过引入蚌线确定下来了。

尼科米迪斯不仅把蚌线用于三等分角问题，而且还利用它解答了倍立方问题。此外，尼科米迪斯还创制了蚌线的机械作图器（下图所示的是 17 世纪的仿制品）。后来，这种曲线在 16、17 世纪又引起许多数学家的兴趣。



化圆为方的历史解答

圆和正方形都是常见的图形，怎样用尺规作一个正方形与已知圆等面积？在历史上，也许没有任何一个几何问题像这个“化圆为方”问题那样强烈引起人们的兴趣。早在公元前 5 世纪，就有许多人在研究这个问题。古希腊著名哲学家阿那克萨戈拉（约公元前 500—前 428）就是其中著名的一位。

阿那克萨戈拉出身显贵，非常富有。但他漠视金钱，将继承的遗产分赠亲属，自己专心于学业。有人问他活着的目的是什么？他回答是为了研究太阳、月亮和天体。他曾主张“太阳是炽烧的石头”，根本不是什么“阿波罗神”。后来，他因这种亵渎神明的观点而被投入监狱。为了打发铁窗岁月，他在狱中画圆刻方，思考化圆为方问题。

但他如何研究化圆为方没有留下任何记录。

在公元前 5 世纪，对这一问题的研究已是如此普遍，以至于古希腊人对那些“献身于化圆为方问题”的人用一个专门的词来表示。古希腊著名剧作家阿里斯多芬的喜剧《鸟》（公元前 414 年上演）中，有一位角色说了这么一段台词：“如果我将这个尺子从上面放下去，再这样插进一个圆规，你看见吗，如此这般，我就化圆为方了。”可见当时化圆为方已经流行，而且自称能够解决这一问题的人已成为被嘲讽的对象。但在这众多的探索中，也确实有不少研究者给出了有价值的研究成果。下面我们就介绍其中几种化圆为方问题的解答。

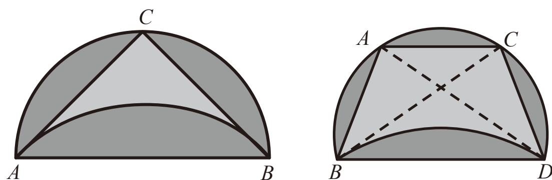
希波克拉底月形

在化圆为方问题方面，一项有名的成果是由古希腊著名数学家希波克拉底（约公元前 470—前 410）获得的。

希波克拉底本是一个商人，但他的船被海盗劫走，他的财产丧失殆尽。他的人生轨迹因此完全改变。为诉讼和查访，他来到了雅典，在那里逗留了很长时间。期间，他常到学校听课，并转向几何学研究，后来他为几何学的发展做出了重要贡献。他曾完成第一本几何教科书，而欧几里得的《几何原本》就是以它为蓝本的。他还致力于“倍立方”与“化圆为方”问题的研究，在前面我们已经介绍了他在倍立方问题上的成果，在化圆为方方面，他成功地将几种月形化方。下面我们就来看一下他在这方面的研究。

我们先来介绍一下希波克拉底考虑的第一种月形。如下面的左图所示，$\triangle ABC$ 为等腰直角三角形。以 AB 为直径的半圆与 AC、BC 形成两个弓形（弓形是指一个圆内被一弦与所截弧所夹的区域）。再以 AB 为底作一个弓形，并使弓形的圆弧与 AC、BC 都相切。于是，$\overset{\frown}{AB}$、$\overset{\frown}{AC}$、$\overset{\frown}{BC}$ 所含的圆心角均为 90°。由于弧所含角相等时弓形相似，所以以 AB 为底所作的大弓形与两个小弓形相似。再进一步，可以证明相似弓形的面积与底长平方成正比。于是根据 $AB^2=AC^2+BC^2$，得到弓形 AB 的面积 = 弓形 AC 的面积 + 弓形 BC 的面积。在这一等式两边都加上"弓形 AB 之上三角形的一部分"，得到两个圆弧所围月形 ACB 的面积等于 $\triangle ABC$ 的面积。

这样，希波克拉底就成功地将曲边图形的面积（月形面积）用直边图形面积（直角三角形面积）表示出来了。



下面再来简单看一下他所考虑的第二种月形。如上面的右图所示，可用尺规画一个等腰梯形 $ABDC$，使其满足 $BA=AC=CD$，且下底 $BD=\sqrt{3}AB$。梯形的外接圆与 BA、AC、CD 形成 3 个弓形。再以 BD 为底作一弓形与 3 个小弓形相似。于是由 $BD^2=3AB^2$，得到弓形 BD 的面积 =3 个小弓形面积之和。这样，两圆弧所夹月形 $BACD$= 梯形的面积。于是，这种月形也可以化方。

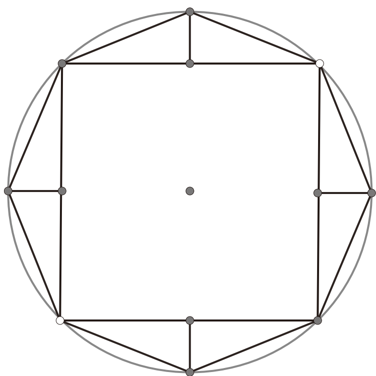
我们看到，在第一种情形中，希波克拉底将一个外圆周等于半圆的月形化为等腰直角三角形。在第二种情形中，他又将一个外圆周大于半圆的月形化为等腰梯形。无论是等腰直角三角形还是等腰梯形，它们都可以容易地化为与之等面积的正方形，因此希波克拉底成功地将上述两种月形化方了。此外，当外圆周小于半圆时，希波克拉底也应用基本作图解决了月形化方。

于是，在外圆周等于半圆、外圆周大于半圆、外圆周小于半圆这三种情形下，希波克拉底都成功地化月形为方。但他并没有将所有的月形都化方。事实上，可以证明有些月形是无法用尺规化方的。

希波克拉底的成果朝解决化圆为方问题的方向迈进了一步。他证明了一系列特殊月形可化方，但对于单个圆的化圆为方，他最终未能解决。虽然如此，这位在几何学上取得卓越成果的数学家仍然赢得了后人的赞誉：“据说最先有效地简化这些困难作图的是希俄斯的希波克拉底，他还化月形为方，并作出许多几何学上的其他发现。说到作图，如果曾经有过这方面的天才的话，这个人就是希波克拉底。”

穷竭法与化圆为方

为了解决化圆为方，智人学派的代表人物安蒂丰（约公元前480—前411）首创了颇有价值的“穷竭法”，提出用圆内接正多边形逼近圆的方法来化圆为方。其具体想法是，画一个圆，并作一个内接正方形；将正方形的每边等分成两部分，从分点向圆周作垂线，显然这些垂线平分圆周上的相应弧段；接着从垂线与圆周的交点向正方形边的端点连线，于是得到4个以线段（即正方形的边）为底的等腰三角形，整个内接的图形现在成为正八边形。以同样的方法重复这一过程，于是可得到内接正16边形、正32边形……随着圆面积逐渐穷竭，一个多边形将内接于圆，由于其边极微小，将与圆重合。



《几何原本》中早已证明能作出一个面积等于任何已知多边形的正方形，那么注意到与圆重合的多边形与圆等积，事实上我们就作出了面积等于一个圆的正方形。

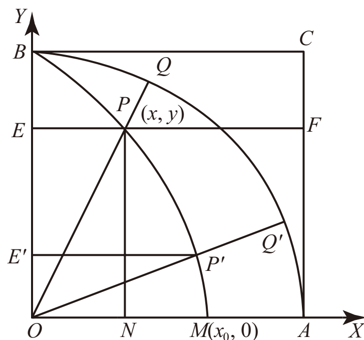
这种推理当然没有真正解决化圆为方问题，但这种有价值的想法成为古希腊“穷竭法”的始祖。在安蒂丰提出这种方法后，古希腊著名数学家欧多克索斯对其进行了改进，使其严格化，完善、成熟的穷竭法成为古希腊人重要而得力的证明方法，并成为后世发现微积分的出发点。事实上，微积分正是基于穷竭法产生的。

割圆曲线与化圆为方

为了化圆为方，古希腊数学家狄诺斯特拉托斯（约公元前 390—前 320）还使用了一种具有特殊性质的曲线：割圆曲线。

割圆曲线是通过运动的合成来定义的。如下图所示，设 *OACB* 为正方形。让 *OB* 以顺时针方向绕 *O* 点以固定的角速度匀速转动，同时让 *BC* 平行于自身以固定的线速

度匀速向下移动，两种运动同时开始并在同一时刻到达 OA 的位置。若用 OQ 和 EF 分别表示这两条运动线段在任一时刻的位置，则它们的交点 P 所形成的轨迹便是割圆曲线。



按割圆曲线的定义，有 $\angle AOQ:\angle AOB=\overset{\frown}{AQ}:\overset{\frown}{AB}=PN:OB$。以 O 为原点，OA 为横轴、OB 为纵轴建立直角坐标系，设正方形的边长为 1，动点 P 的坐标为 (x, y)，于是有：$\dfrac{\angle AOP}{\angle AOB}=\dfrac{y}{1}$。设 $\angle AOP=\theta$，则 $\dfrac{\theta}{\frac{\pi}{2}}=\dfrac{y}{1}$，所以 $\theta=\dfrac{\pi}{2}y$，又 $\tan\theta=\dfrac{y}{x}$，因此，$x=\dfrac{y}{\tan\theta}=\dfrac{y}{\tan\frac{\pi y}{2}}$。

设曲线与 OA 的交点 M 的坐标为 $(x_0, 0)$，利用极限知识可得：

$$x_0=\lim_{y\to 0}x=\lim_{y\to 0}\frac{y}{\tan\frac{\pi y}{2}}=\lim_{y\to 0}\frac{y}{\frac{\pi y}{2}}=\frac{2}{\pi}。$$

这一等式表明，如果能作出割圆曲线，则能作出线段 $\dfrac{2}{\pi}$。通过比例作图进而获得线段 $\dfrac{\pi}{2}$，而以线段 $\dfrac{\pi}{2}$ 为一边、2 为另一边的矩形可以化为等面积的正方形，其面积恰为单位圆的面积。因此，借助割圆曲线可以化圆为方。事实上，早在公元前 350 年，狄诺斯特拉托斯就已经成功地用割圆曲线完成了化圆为方。不过，在关键的 OM 长度的确定上，他是用观察法确定出其值为 $\dfrac{2}{\pi}$，然后又用古希腊人擅长的严

格穷竭法证明了这一结果。因为割圆曲线可用以化圆为方，所以它也被命名为圆积曲线。

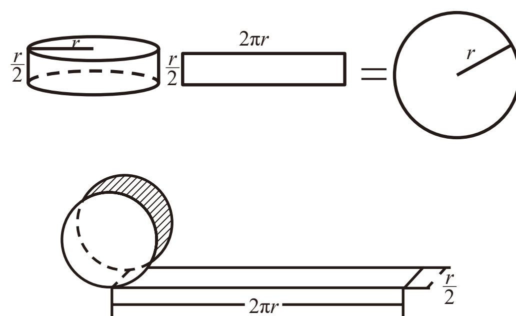
顺便指出，割圆曲线还可以用来三等分角。如上图所示，设 $\angle AOQ$ 是要三等分的角，OQ 交割圆曲线于 P，作 $PE \perp OB$。为了三等分 $\angle AOQ$，只需三等分对应的线段 OE。取 $OE' = \frac{1}{3}OE$，然后作 $E'P' \perp OB$ 交割圆曲线于 P'，则 $\angle AOP'$ 即为所求的三等分角。理由很简单，根据割圆曲线的特殊性质：$\angle AOP : \angle AOB = PN : OB = EO : OB$ 与 $\angle AOP' : \angle AOB = E'O : OB$，于是 $\angle AOP' : \angle AOP = E'O : EO = 1:3$。显然用这种方法还可以轻易地解决 n 等分任意角问题，因为只需 n 等分对应的线段即可。事实上，正是为了解决三等分角问题，大约在公元前 430 年，古希腊数学家希比亚斯（约公元前 460—前 400）首先发现了这种除直线和圆以外的曲线。

当然，虽然能用割圆曲线三等分角与化圆为方，但是这种曲线本身是无法通过尺规作出的。

达·芬奇做法

文艺复兴时期，著名画家达·芬奇曾给出化圆为方的一个简易而巧妙的方法：设已知圆的半径为 r，则其面积为 πr^2。作一底面半径为 r、高为 $\frac{r}{2}$ 的圆柱，将此圆柱在平面上滚动一圈，得到一矩形。这一矩形的长为底面圆的周长 $2\pi r$，宽为 $\frac{r}{2}$，其面积为 $2\pi r \cdot \frac{r}{2} = \pi r^2$，恰为已知圆的面积。剩下的问题是求作一个与此矩形面积相等的正方形，这是非常容易做到的。这样就完成了化圆为方。当然，达·芬奇做法也不符合尺

规作图的要求。





第三节
不可解的证明

几何三大作图问题历经上千年，耗费了无数数学爱好者与数学家的心血，却始终未获真正的解决。为什么貌似简单的三大几何作图问题一直不能求解？直到 19 世纪，在经过 2000 多年后，谜底终于揭开了：用尺规完成三大作图问题“非不为也，是不能也”。换句话说，这三大作图问题是尺规不能问题。谜底是如何揭开的呢？我们要从搭起几何与代数之桥的解析几何说起。

解析几何的建立

数学最早起源于古埃及、古巴比伦、古中国、古印度四大河谷文明。从公元前 6 世纪开始，古希腊人从古埃及、古巴比伦人那里学习数学，并对其进行改造，特别是引入了证明的思想，使数学的本质发生了改变。青出于蓝的古希腊成为当时世界数学的发展中心。在古希腊人手中，传统几何学被发展到非常完善的程度。但随着古希腊数学帷幕的落下，欧洲数学

进入一个漫长的黑暗时代，数学的中心重新转移到了东方，中国、印度与阿拉伯成为数学活跃的舞台。在这一阶段，代数作为一门独立的学科逐渐得到确立。

1100年左右，贸易、旅游、战争等因素促成了东西方文化的交流。随着东方文化的传入，欧洲开始从黑暗时代觉醒了，欧洲数学也开始复苏，并开始逐步进入发展的快行道。特别是在15、16世纪，西方数学首先在代数方面取得了重大突破。一方面，代数方程论取得了长足的进展，意大利数学家获得了三次、四次方程的公式解法；另一方面，人们在代数中引入符号，特别是法国著名数学家韦达完成第一部符号代数著作《分析术引论》，为代数符号体系的建立做出了卓越贡献。在韦达之后，数学符号仍被陆续引入，符号体系不断得到改进和完善，最终形成了我们现在所使用的简洁、优美的数学符号体系。随着代数方法和符号的充分发展，代数这一学科日趋成熟，越来越成为解决问题的有效工具，其威力随着它的成熟日渐显现。于是，通过代数解决几何问题，找到研究曲线的新途径成为数学发展的趋势。两位伟大的法国数学家笛卡儿、费马都在对圆锥曲线的研究中，成功创建了解析几何，并共享了创建解析几何带来的荣誉。关于费马，我们会在后面章节中做介绍，这里先来了解一下笛卡儿（1596—1650）。



笛卡儿

笛卡儿，1596 年 3 月 31 日出生在法国一个古老的贵族家庭。自幼年起，被父亲称为“小哲学家”的笛卡儿就对世界充满好奇，总想知道阳光下万物的本原。由于体弱多病，直到 8 岁，笛卡儿才被送进一所欧洲著名的耶稣会学校接受正规教育。得到校方的允许后，笛卡儿被特许每天在床上多睡些时间。由此，他养成了早晨沉思的习惯。中年时回顾自己这段学生生活，笛卡儿认为“那些在寂静的冥思中度过的漫长而安静的早晨，是我的哲学和数学思想的真正源泉”。

1616 年，笛卡儿被授予法学博士学位。此后不久，他开始到社会上去阅读“世界这本大书”。1617 年 5 月，笛卡儿选择了参军。在度过几年军队生活后，1621 年他离开军队，随后到丹麦、荷兰、瑞士、意大利等地游历。1625 年他回到巴黎，与梅森、笛沙格等共同研究数学，同时进行其他方面的探索。从 1628 年到 1649 年，笛卡儿定居荷兰，并将自己的思想加以总结，潜心著述，完成了他所有的主要著作。

笛卡儿一生涉猎极广。他最具影响的方面是哲学与数学。在哲学上，黑格尔曾称他为“现代哲学之父”。在数学上，他贡献很多，比如他对符号系统改革做出的重大贡献，其中一项是用字母表中开头几个字母 a、b、c 等来表示已知数，用末尾几个字母 x、y、z 等表示未知数。他最重要的发现是创建解析几何这一影响深远的数学研究领域。正如许多伟大发现常会伴随美妙的传说一样，关于笛卡儿创立解析几何的灵感也有一则带有神秘色彩的佳话。

1619 年冬天，笛卡儿随军队驻扎在多瑙河畔的一个村庄。11 月 10 日，圣马丁之夜。在一场庆祝圣马丁节的盛宴之后，笛卡儿连续作了三个奇特的梦。他先是梦到自己被风暴从安全居所席卷到风力无法摇撼的地方。在随后的梦境中，他发现自己观察着凶猛的风暴，并意识到自己一旦看出风暴是怎么一回事，风暴就不能伤害到他了。梦中的他忽然获得一种顿悟，似乎梦境向他揭示了一把可以打开大自然宝库的魔匙，使他能够掌握所有科学的真正基础。梦境再转时，梦中的他朗诵起奥索尼厄斯的诗句：“我将遵循什么样的生活道路?”

笛卡儿后来说正是这三个梦向他揭示了“一门奇特的科学”和“一项惊人的发现”，虽然他从未明说过这门奇特的科学和这项惊人的发现是什么，但人们通常推测他在梦境中所获得的那把能打开大自然宝库的魔匙就是解析几何。即便接受这种推测，我们仍然可以确信笛卡儿在梦境中获得的至多只是一种启示，而要真正把握住这把魔匙，他还需要付出更多的探索与努力。而人类的智慧库中添加上这把魔匙还需要再等待 18 年的岁月。

1637 年 6 月 8 日，笛卡儿出版《更好地指导推理和寻求科学真理的方法论》(简称《方法论》) 一书，作为他提出的一般方法论的应用，他在书后加有三个附录。在作为附录之一出现的“几何学”中，笛卡儿阐述了他的解析几何思想。《方法论》出版的这一天也就成为解析几何诞生的日子。

解析几何的创建“构成了人类在精确科学的进步上曾迈出的最伟大一步”，它使数学的面貌焕然一新，成为数学发展中具有深远意义的大事。它的建立跨出了从常量数学到变量数学的第一步；它直接促成了微积分的诞生；它把以前被人为割裂的代数与几何、数与形都重新黏合在一起。它的出现使数学领域内的代数、几何这两个不同学科互相渗透，从而为数学提供了一个双面的工具。几何概念可用代数表示，几何的目标可通过代数达到。反过来，人们给代数语言以几何的解释，就可以直观地掌握那些语言的意义，又可以得到启发去提出新的结论。

对此有深刻领悟的笛卡儿在仅 117 页的篇幅中，清楚证明了几何问题可以归结为代数形式的问题，因此在求解时可以运用代数的全部方法。作为这种新思想的应用，笛卡儿在题为“关于只用圆和直线的作图可能问题”的一卷中就把几何问题化成代数问题，提出了几何问题的统一作图法。因此，对于本章要探讨的主题而言，解析几何的重要性在于为三大几何问题的研究开辟了一条新的道路。

尺规的能力

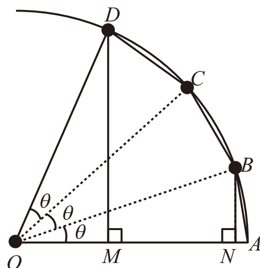
解析几何的创立，使几何问题转化为代数问题成为可能。借助解析几何思想，人们可以把作图问题转化为对应的代数问题。

正如我们已经介绍的，尺规作图就是在已知的一些初等平面几何图形（点、直线、圆）的基础上，利用直尺和圆规，遵循作图公法，作出新的图形。做进一步分析：直线由其上两点决定，两点只要确定了，按照作图公法1°，则这条直线可以作出；而圆由圆心和半径决定，或者说由圆心和圆上一点决定，只要它们确定了，按照作图公法2°，则圆可以作出。因此，直线与圆的作图归结为点的确定。如果我们在平面上引入直角坐标系，即指定原点和坐标轴，并指定单位长度，那么因为直线、圆归结为点，点又归结为坐标（即实数），这样作图问题就转化为，已知一些实数，在此基础上用尺规作图得到所求的实数。由于数与线段（长）的对应，作图问题也可以表述为，给定某些线段，求作一个或多个其他线段。

于是，任何几何作图可归结为一个代数问题：首先，我们需要找出所求的量和给定的量之间的关系（方程）；然后，解方程求未知量；最后需要确定，通过用尺规来作图的代数过程能否得到这个问题的解。

下面我们以三等分角问题为例说明一下这种由几何作图到代数问题的转化。

如下页图所示，我们打算三等分 $\angle AOD$。以 O 为圆心作单位圆，它与角的两边相交于 A、D，则 $OA=OD=1$。设 $\angle AOD$ 的两条三等分角线与弧 AD 相交于 B、C，显然，角的三等分问题等价于 B、C 的确定问题。当然，B、C 两点只要确定一个，另一个也就可以确定下来。现在我们考虑点 B 是否可以找到。



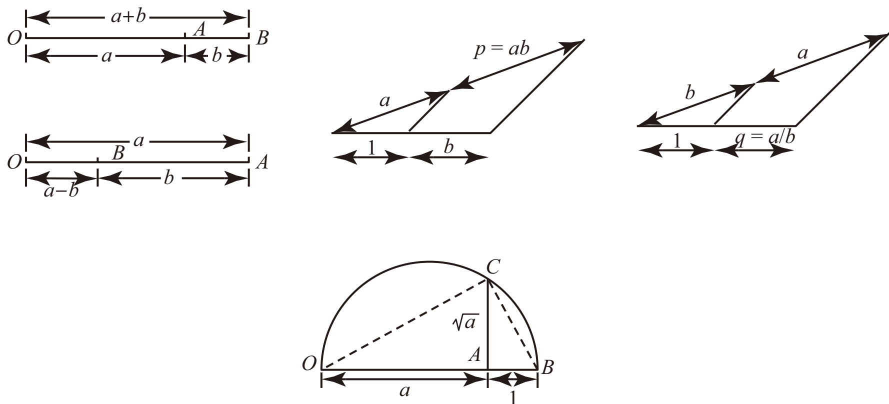
分别过 D、B 作 OA 的垂线，它们与 OA 相交于 M、N，于是找 B 点的问题转化为确定点 N 的问题。

设 $\angle AOD=3\theta$，则 $\angle AOB=\frac{1}{3}\angle AOD=\theta$，于是 $OM=\cos 3\theta$，$ON=\cos\theta$。$\angle AOD$ 已知，则 OM 已知，不妨设 OM 为 a。而用尺规确定点 N，只要作出线段 ON，或者说确定线段 ON 的长。不妨设 ON 为 x。这样，三等分角这一几何作图问题就转化为由已知的量 OM（a），求未知的量 ON（x）。

下一步则是建立两者之间的关系（方程）。由三倍角余弦公式 $\cos 3\theta=4\cos^3\theta-3\cos\theta$，可得 $a=4x^3-3x$，即 $4x^3-3x-a=0$。

为了确定通过用尺规来作图的代数过程能否得到这个问题的解，我们还需要先从代数角度考察一下直尺与圆规究竟有多大的能耐。

首先，我们来看初等几何作图的一些最简单的代数运算。如下页图所示，若给定两个长为 a 和 b 的线段，可作出 $a+b$、$a-b$、ab、$\frac{a}{b}$。由此可知，已知量的加减乘除这些有理运算能用几何作图来实现。连续应用这种简单作图，我们就能从长度为实数 $a, b, c, \cdots$ 的任意给定线段出发，作出用 $a, b, c, \cdots$ 的有理式来表示的任意量。尺规作图的能力还不限于此，它还能作出平方根。如图，给定一条线段 a，则用尺规可作出 $\sqrt{a}$。连续应用这种简单作图，还可作出 $\sqrt[4]{a}, \sqrt[8]{a}, \cdots$。当然，以上简单作图还可以进行任意有限次组合。于是，由已知数 $a, b, c, \cdots$ 有限次加减乘除及开平方运算得到的所有数都可以用尺规作出。



对以上结论，我们下面以一种简单的情况为例再说明一下。

假定最初只给定单位长度 1，通过加减乘除这四种有理运算可以得到所有有理数。利用尺规，还能作出新的无理数，比如说 $\sqrt{2}$ 。在作出 $\sqrt{2}$ 后，又可进而作出所有形如 $a+b\sqrt{2}$（a, b 是有理数）的数。这一过程可反复进行有限次。因此，只要有理数经过有限次“加减乘除开平方”五种运算得出的数，都可以用尺规作出。这样的数（线段）也称为“可构造数”或“可作图量”。比如，单位圆的内接正五边形的边长 $\frac{5}{8}\sqrt{10+2\sqrt{5}}$ 一定可以用尺规作出，这也意味着，用尺规作圆的内接正五边形是可能的。

如果考虑从长度为实数 a, b, c, … 的任意给定线段出发，则结论基本类似：由给定数（线段）经过有限次“加减乘除开平方”运算得出的数（线段），都可以用尺规作出。

如果用上一章介绍过的“域”的概念，我们可以换一种方式表述这一结果，即从给定的数 1, a, b, c, … 出发，以“加减乘除”得到的所有量构成了一个数域 F ，这个域是由给定的数 1, a, b, c, … 构成的（特别地，如果给定数只有 1，则得到有理数域 Q ）。由数域 F 出发，加入“开平方”运算，可以得到 F 的一阶二次扩域 F_1 中的所有数（线段）。在此基础上，还可以继续扩充，得出 F 的二阶二次扩域（即 F 的 4 次扩域）、

三阶二次扩域（即 F 的 8 次扩域）…… F 的 n 阶二次扩域（即 F 的 2^n 次扩域）中的所有数（线段）。F 及 F 的这一系列扩域中的数（线段）都是尺规可作的。

至此，我们清楚了尺规作图能做什么。下面我们要探讨的是，尺规作图能做的是否仅限于此。

如我们所知，尺规作图的过程无非是从一些已知点（数）出发来求一些新的点（数）。而根据作图公设 3°，这些新的点只能通过求直线和直线、直线和圆、圆和圆的交点来得到。在建立直角坐标系后，直线和圆的一般方程可分别表示为 $Ax+By+C=0$ 与 $x^2+y^2+Dx+Ey+F=0$，方程中的系数自然属于由已知数构成的数域 F。另外需要说明的是，在尺规作图的过程中，经常会出现"任意"的元素，如任取一点、画任意一条直线、画任意一个圆，在这种情况下，因为元素是任意的，我们自然可以选取属于由已知数构成的数域 F 的数（如点的坐标、直线或圆的方程的系数）。下面考虑几种交点的情况。

先考虑两直线的交点，其坐标由下列方程组确定：

$$\begin{cases} A_1x+B_1y+C_1=0 \\ A_2x+B_2y+C_2=0 \end{cases}$$

容易知道，交点坐标可由系数经过有限次的加减乘除求得。因此，在这种情况下，得到的数一定不超出数域 F 的范围。

再考虑直线与圆的交点，其坐标由下列方程组确定：

$$\begin{cases} Ax+By+C=0 \\ x^2+y^2+Dx+Ey+F=0 \end{cases}$$

由代数知识可知，这一方程组的解可以由直线或圆方程中的系数经过有限次的加减乘除和开平方求得。因此，在这种情况下，得到的坐标（数）一定不超出数域 F 的某阶二次扩域。

最后考虑两圆的交点，其坐标由下列方程组确定：

$$\begin{cases} x^2+y^2+D_1x+E_1y+F_1=0 \\ x^2+y^2+D_2x+E_2y+F_2=0 \end{cases}$$

这一方程组等价于解方程组

$$\begin{cases} (D_1-D_2)x+(E_1-E_2)y+(F_1-F_2)=0 \\ x^2+y^2+D_2x+E_2y+F_2=0 \end{cases}$$

与上一种情况相同，这一方程组的解可以由系数经过有限次的加减乘除和开平方求得，得到的坐标（数）一定不超出数域F的某阶二次扩域。

综合以上的讨论，我们看到，如果某一数（线段）可以由已知的数（线段）经有限次的加减乘除及开平方后得出，则此数（线段）一定可以用圆规及直尺作出来。而尺规的本领也仅限于此。用"域"的概念表述，即尺规可作出由已知数构成的域及此域的2次、4次、8次……扩张域中的数。

也就是说，尺规作图的本领有限，其能耐仅此而已。由此，我们也有了一个判定某作图问题是尺规作图可能问题还是尺规作图不能问题的准则，即某几何作图所求数（线段）当且仅当可由已知数（线段）经有限次四则运算及开平方运算得出，或者说所求数是已知数构成的域及某阶二次扩域中的数时，可用尺规作图。反之，如果某几何作图所求数（线段）不能从已给数（线段）经有限次四则运算及开平方运算得出，或者所求数不是已知数构成的域及某阶二次扩域中的数，则此作图是尺规作图不能问题。

有了这一判定准则后，我们可以去解释2000多年以来人们一直未能用尺规完成三大问题的原因了。

三大问题的解决

有了上面的准备后，我们先来研究一下倍立方问题。首先，倍立方问题可转化为一个代数方程求解问题。

设已知立方体的棱长为 a，倍立方问题就是要作出新立方体，其棱长为 x，使得 $x^3=2a^3$，即要求作出 $x=\sqrt[3]{2}a$。因为只要能作出 $\sqrt[3]{2}$，即可利用尺规容易地作出 $\sqrt[3]{2}a$，因此倍立方问题可简化为考察 $x=\sqrt[3]{2}$ 或者说方程 $x^3-2=0$ 的实解 x 是否可用尺规作出。从直觉上看，由有理数域 Q 出发，经过有限次开平方运算，不可能得出立方根 $\sqrt[3]{2}$。但要证明这一点并不容易。下面我们用反证法。假定 x 可用尺规作出，并打算由此假定推出矛盾。

根据上面已经介绍到的，如果 x 能用尺规作出，那么 x 必属于有理数域 Q 或 Q 的某扩域，这个扩域由 Q 经有限次加法运算及开平方得出。就我们要探讨的这一问题而言，因为 $\sqrt[3]{2}$ 不是有理数，因此 x 必属于有理数域 Q 的某一扩域 F_k，显然我们可以使 F_k 为包含 x 的最小扩域。这意味着，x 可表示成 $a+b\sqrt{d}$ 的形式，其中 $a,\ b,\ d\in F_{k-1}$，但 $\sqrt{d}\notin F_{k-1}$。

下一步我们要说明的是，若 $x=a+b\sqrt{d}$ 是 $x^3-2=0$ 的一个根，那么 $a-b\sqrt{d}$ 也是 $x^3-2=0$ 的一个根，且此根为实数解。

我们把 $x=a+b\sqrt{d}$ 代入方程 $x^3-2=0$ 中，得到 $(a+b\sqrt{d})^3-2=0$。展开，得 $(a^3+3ab^2d-2)+\sqrt{d}(3a^2b+b^3d)=0$。如果 $3a^2b+b^3d\neq 0$，那么可推出 $\sqrt{d}=-\dfrac{a^3+3ab^2d-2}{3a^2b+b^3d}$，于是 $\sqrt{d}\in F_{k-1}$ 与已知 $\sqrt{d}\notin F_{k-1}$ 矛盾。所以 $3a^2b+b^3d=0$，由此马上又可得出 $a^3+3ab^2d-2=0$。于是又有 $(a^3+3ab^2d-2)-\sqrt{d}(3a^2b+b^3d)=0$，即

$(a-b\sqrt{d})^3-2=0$。这说明$a-b\sqrt{d}$也是$x^3-2=0$的一个根。

根据图像可知，$y=x^3-2$与x轴只有一个交点，即$x^3-2=0$只有一个实数解。而现在我们得到了$a+b\sqrt{d}$与$a-b\sqrt{d}$都是方程$x^3-2=0$的实解，于是矛盾。这说明我们最初的假定错误，因此，x不属于有理数域Q的任一扩域。这表明用尺规无法完成倍立方问题。

上述证明思路还可以用来证明一个关于三次方程的一般结论：有理系数的三次方程$x^3+bx^2+cx+d=0$如果没有有理根，那么它的每一个根都不能用尺规作出。

仍然用反证法。假设这样的方程有根可用尺规作出，则这根必属于某个域F_k（显然可以使F_k为包含x的最小的扩域），这里$k\geqslant 1$，$F_0=Q$。于是这根可写成$g+h\sqrt{a}$的形式，其中$g,h\in F_{k-1}$，$\sqrt{a}\notin F_{k-1}$。下一步与上面的方法类似，通过将$g+h\sqrt{a}$代入方程，展开考虑不含$\sqrt{a}$的“有理部分”与含$\sqrt{a}$的“无理部分”。最终推得$g-h\sqrt{a}$也是此方程的根。这样，根据韦达定理，此方程的第三个根$x''+(g+h\sqrt{a})+(g-h\sqrt{a})=-b$，于是$x''=-b-2g\in F_{k-1}$，这与$F_k$为包含$x$的最小扩域矛盾。于是假设错误，命题得证。

有了这个一般结论后，我们再来看一下三等分角问题。

在上面的介绍中，我们已经提到三等分角问题可以归结为考察方程$4x^3-3x-a=0$。给定的已知角不同，则已知数a不同。比如，给定的已知角为$90°$，则$a=0$，于是上述方程转化为$4x^3-3x=0$，可解得正根$x=\dfrac{\sqrt{3}}{2}$，这是尺规可作量，因此直角能用尺规三等分。这一特例表明，有些角度是可以用尺规三等分的。现在，我们来看另一个特例，即取已知角为$60°$，则$a=\dfrac{1}{2}$，于是上述方程化为$4x^3-3x-\dfrac{1}{2}=0$，即$8x^3-6x-1=0$，那么此方程的解是否为尺规可作呢?

在方程理论中，有这样一条结论：如果一个多项式方程$a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n=0$有有理根$\dfrac{a}{b}$，则$a$一定是方程中常数项$a_n$的因子，$b$一定是方程中最高次项系数$a_0$的因子。应用这一结果，$8x^3-6x-1=0$若有有理根，也只可能是$\pm 1$，$\pm\dfrac{1}{2}$，$\pm\dfrac{1}{4}$，$\pm\dfrac{1}{8}$，把

它们一一代入方程，发现都不成立。因此，$8x^3-6x-1=0$ 没有有理根。再根据上面已经证明的关于三次方程的结论，即可得出此方程确定的解不能用尺规作出。这说明，无法用尺规三等分 60° 角。这一特例表明，用尺规三等分任意角是不可能的。

由此，我们给出了倍立方问题与三等分角问题的最终答案：它们都是尺规不能问题。问题通过否定的方式获得了解决。通过上述介绍，我们还可注意到，这两个问题的解决除了要借助由解析几何知识得到的可否尺规作图的判定准则之外，还要应用方程论知识及域论的思想。正因此，这两大问题的最终解决，在解析几何诞生之后还要等待整整两个世纪。1837 年，法国数学家旺策尔（1814—1848）研究阿贝尔定理的化简时，在代数方程论的基础上首先严格证明了倍立方和三等分任意角这两个问题不能用尺规作图完成。

下面我们再来说明一下化圆为方问题是尺规作图不能问题。为此，我们需要先引入代数数的定义。

如果一个数满足下面的代数方程 $a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n=0$，其中方程中的系数都是整数，那么这个数叫作代数数。如果它又是实数，则称为实代数数。不是代数数的数叫作超越数。对代数数还有进一步的定义：若 x 满足的最低的代数方程的次数是 n，则 x 叫作 n 次代数数。根据这种定义，可以知道有理数都是一次代数数。$\sqrt{2}$ 是代数数，并且是 2 次代数数，因为它满足的最低次代数方程是 $x^2-2=0$。再如，$\sqrt[3]{2}$ 是代数数，并且是一个 3 次代数数，因为它满足的最低次代数方程是 $x^3-2=0$。

关于尺规可作的数，如果用代数数的语言表述，则有如下结论：尺规可作的数仅限于一次代数数、2 次代数数、4 次代数数…… 2^n 次代数数，除此之外的数都不能通过尺规作出。再举一个例子说明一下，比如 $x=\sqrt{2}+\sqrt{3+\sqrt{2}}$，它满足一个整系数的四次方程：

$$x^2+2-2\sqrt{2}x=3+\sqrt{2}$$

即 $x^2-1=\sqrt{2}(2x+1)$，两端再平方得 $(x^2-1)^2=2(2x+1)^2$，显然这是一个整系数的四次方程。这表明，$x=\sqrt{2}+\sqrt{3+\sqrt{2}}$ 是一个 4 次代数数，因此它是尺规可作的。

显然，尺规可作的数只是代数数中很小的一部分。这也意味着，如果一个数根本不是代数数，那么它一定是无法用尺规作出的。由此，我们可以去看一下化圆为方问题的解答了。

首先，化圆为方问题可转化为一个代数方程求解问题。假设已知圆的半径为 r，求作的正方形的边长为 x。按条件，应有 $x^2=\pi r^2$，即 $x=\sqrt{\pi}r$。只要能作出 π，即可利用尺规容易地作出 $\sqrt{\pi}$，进而利用尺规容易地作出 $\sqrt{\pi}r$，因此化圆为方问题可简化为考察在给定单位长度 1 后，π 是否可作。

1761 年，数学家兰伯特证明 π 是无理数。但这一结论仍无法确定 π 是否尺规可作。有数学家进而猜测 π 不可能是代数数，它既然连代数数都不是，就更不是 2 次或 2^k 次代数数。但证明 π 是超越数是非常困难的。1873 年，法国数学家埃尔米特第一次证明了一个数学中的常见数 e 是超越数，证明过程很不简单。这一次成功，使得许多人打算用埃尔米特的方法去证明圆周率 π 的超越性。埃尔米特听到这事后说："我可以不再冒险去尝试证明 π 的超越性。假如别人打算干，我会比别人更高兴看到他们取得成功。但是，相信我，取得成功绝不是轻而易举的事。"

然而，事情的发展出人意料，1882 年德国数学家林德曼基本上用埃尔米特的方法证明了 π 是超越数。为此，林德曼被人们称为"π 的战胜者"。这一结果同时使得化圆为方问题获得了圆满解决。结论是，用尺规解决化圆为方问题也是不可能的！林德曼等人的研究还开拓了一个研究超越数的理论。这一新的数学领域叫作超越数论，它构成了数论的一个分支。

历时 2000 多年的几何三大问题最终被人类征服了。1895 年，德国数学家克莱因在总结前人研究成果的基础上，在德国数理教学改进社开会时宣读了一篇论文。论文

中他又给出几何三大问题不可能用尺规作图的简单明晰的证法，从而使 2000 多年以来未解决的疑问告一段落。

结　　语

前面我们大致介绍了几何三大问题由提出到解决的历史，在离开这一站之前，让我们稍作逗留，回顾一下这一段行程并做几点思考。

三大作图问题自公元前 5 世纪由智人学派提出之后，在 2000 多年的时间里，以其历久不衰的魅力吸引了无数人的注意与研究，其中包括一代又一代最为优秀的数学家。这三大问题的妙处在于它们看上去非常简单，但由于限制作图工具只能用尺规，这些问题变得难以解决并更富挑战性。在问题提出后，人们最初千方百计寻求问题的正面解答。沿着这一方向辛勤努力的结果是，大量不限尺规的解答被提出。除我们在本书中介绍的做法外，还有众多的解法我们没有提到。比如，单是古希腊学者就留下了倍立方问题的 12 种解答，其中较有名的是：古希腊数学家阿尔希塔斯（约公元前 428—前 347）用了一种十分漂亮且技巧性很强的方法；古希腊数学家狄奥克勒斯约在公元前 2 世纪末用自己发现的蔓叶线解决了倍立方问题。再比如，三等分角问题有帕普斯的双曲线解法、阿基米德螺线解法、笛卡儿做法、帕斯卡蜗线做法、牛顿做法等。化圆为方问题有阿基米德螺线解法等。然而，所有这些历史解答都因无法严格遵守尺规作图的限制，而未能真正解决三大作图问题。为什么如此多的努力都以失败而告终？是所有这些研究者缺乏足够的耐心和聪明才智还是问题本身不可解？什么样的图

形可以用尺规作出？什么样的图形无法用尺规作出？“不识庐山真面目，只缘身在此山中”，这样的问题是欧几里得几何本身不能解决的。

解析几何的出现，翻开了三大问题研究的新的一页。借助解析几何，人们可以思考尺规作图的能力，并给出尺规可作图与不可作图的判定准则。由此，人们开始怀疑三大问题一直未能获得真正解决的原因在于这样的解法根本不存在。如解析几何创始人笛卡儿就曾指出倍立方问题是不可用尺规完成的。但为了严格证明这一点，还需要等待数学的进一步发展。直到 1837 年，数学家借助方程理论与域论思想才首次严格证明了倍立方与三等分角问题是尺规不能问题。事实上，如果利用后来的伽罗瓦域扩张理论，这两个问题的解答会更为简洁：设 α 是有理系数不可约三次方程的一个根，则可得到扩张次数 $(Q(\alpha):Q)=3$ ，而尺规可作数所属的域扩张次数只能是 2^n ，并且 $3\neq 2^n$ ，又利用方程理论可说明倍立方问题与三等分角问题中出现的方程是不可约三次方程，于是其尺规不可解，得证。化圆为方问题的尺规不可解则需归结为证明 π 的超越性。林德曼的成功，既为化圆为方问题画上了句号，又推动了超越数论的研究。在历经 2000 多年后，三大几何问题终于都以否定的方式获得了解答。

当我们回顾这三大问题的解决历程时，我们要说过去人们所做的努力绝非毫无意义。事实上，三大问题正因为不能用尺规来解决，常常使人闯入新的领域中去。结果，这几个著名问题本身虽然没有得到正面解决，但是为解决这三大问题却孵出许多有价值的金蛋。

其一，古希腊几何沿着欧几里得路线研究的几何对象原本仅限于二维的直线、圆与三维的平面、球。然而为了解决几何三大问题，古希腊人发现了一批圆和直线以外的曲线以及球和平面以外的曲面，这促成了古希腊在更为高等的几何方面的发展。这其中，最重要的是开创了圆锥曲线的研究。而圆锥曲线又在 17 世纪召唤出解析几何这门学科。

其二，在化圆为方研究中，由安蒂丰首创的“穷竭法”被古希腊人进一步发展，成为严格证明许多数学命题的有力工具，到 17 世纪更是开花结果，直接促成了微积分的诞生。

其三，为了解决三大问题，人们提出了尺规作图的判定准则。

其四，三大问题的最终解决，与近代方程理论、代数数理论、超越数理论及域论都有关联，因此对前者的研究促进了后者的发展。

可以说，由三大几何问题孵出的这些金蛋的价值远远高于三大几何问题本身所具有的价值，其意义也比几何三大问题本身的意义深远得多。随着这一公案的了结，孵金蛋的鹅被宰杀了，三大问题作为已圆满解决的数学问题对数学的推动作用失去了，它早已不再是需要人们去继续攻攀的难题。然而，古今中外却都有执迷者继续把精力浪费在这三大几何问题的求解上。

比如，在化圆为方问题已被林德曼解决后，仍有许多人为“化圆为方”走火入魔。数学家德 · 摩根把这些热衷“化圆为方”问题的人称为“方圆病”患者。法国一数学家则称这种人是“害了聪明病”的人，他用幽默的口吻写道：“所有国家的科学院，在和追求解决方圆问题的人们作斗争时发现一个事实，这个病症一般都在春天的时候加剧。”也许并不太出乎人们意料的一件事实是，在西方很多“方圆病”患者甚至出现在受人尊敬的高层人士中，比如患者中曾有一个大学校长、一个华盛顿州下院议员，还有一个美国参议院议员。

另有许多“三等分角家”迷恋于三等分角。在美国，数学杂志社每年总要收到许多三等分角的来信；在报纸上还常有人宣称解决了这个千古难题。而我国著名数学家华罗庚曾在 1952 年一年中就答复了将近二三十封关于三分角的信，最后他不得不写了一篇“三分角问题”的科普文章。但这并没有阻止“三等分角家”的热情。事实上，中国科学院数学研究所在 1966 年以前，每年都要收到许多号称“解决三等分角问题”

的来稿，以至于中国数学会及《数学通报》编辑部不得不在《数学通报》上多次发表声明，劝告人们不要在这个不可能的问题上浪费时间和精力。

类似的事情也发生在许多其他著名数学问题身上，如本书后面章节会介绍到的“费马大定理”“四色问题”，还有本书没提到的“哥德巴赫猜想”等，都拥有众多以解决它们为己任的数学“民科”。因此，我们有必要在这里对这一现象多说几句。

首先，我们可以注意到这些患有“聪明病”的业余研究者总想通过独步古今中外，压倒所有前人的工作而证明自己的聪明。其主要动机可用著名作家王小波一篇故事中的一句话来概括：“证明了费马大定理，就证明了自己是世界上最聪明的人，这种事值得一干。”另外，这些研究者的热情与为此做出的牺牲也往往令人油然而生敬意。然而，当我们看到他们所有的努力注定以失败而告终时，我们不禁为其毫无意义的努力与牺牲而扼腕。之所以如此，一方面是因为他们缺乏自知之明。事实上，许多业余研究者对数学的了解仅限于初等数学或初级的高等数学。另一方面，是因为他们对问题缺乏了解。像“三大几何问题”“费马大定理”“四色猜想”等数学问题表述非常简单，很容易理解，然而其妙处却在于为了解决它们要使用新的数学思想或极为复杂的数学理论。本章介绍的几何三大问题带来的启示之一就是如何对待一个长期未解决的问题，特别是历史长、影响深，得到过一些著名数学家钻研而尚未解决的那些著名问题。这些问题往往不是用通常的方法所能解决的。对于那些至今未解决的许多著名问题，都应采取这样的态度，停留于初等方法是绝不可能解决这些问题的。

可惜与可叹的是，业余研究者往往被问题表述的简单性所迷惑，在对问题究竟难在何处几乎无所知的情况下，在仅仅掌握了初等数学知识的前提下，就想依靠“铁棒磨成针”的毅力与苦战过关。这自然只会白白浪费许多宝贵的精力。

以三大几何问题为例，许多研究者宣称自己解决了三大问题中的一个。实际上，他们并不了解问题难解的关键在工具及作图公设的限制。许多人也分不清不可能与未解决的本质区别，不愿或没有能力去理解问题不可解的原因。他们会强调：“从前认为

不可能的事，现在有些是可能了！所以不可能的说法仅仅是阻碍进步，我们不可为不可能三字而限制了我们的前进，限制了我们的发明……所以用圆规及直尺三等分角的问题，也并不是绝对不可能的。”华罗庚先生为此打了一个譬喻：“上月亮去”是一个“未解决”的问题。但“步行上月亮去”是一个“不可能”的问题。用圆规及直尺解决三大几何问题正如步行上月亮一样是不可能的，而不是未解决的。

美籍华裔几何学家杨忠道教授也打过一则很有说服力的比喻。一位在农村成长而未受过任何教育的父亲，非常疼爱他的一位聪明的儿子。在儿子离家几十里外求学的时候，父亲因为太想念儿子，每隔几个月要去探望一次。也许是他个人的习惯，每次去都是步行的。后来儿子去邻县工作，父亲还是维持以往的习惯，每隔几个月就步行去探望儿子一次。最后难题来了，儿子到一海岛上去工作。当父亲想再步行去探望儿子时，一些人告诉他说，那是不可能的。因为父亲没有受过任何教育，他很为下列诸问题所困扰。

(1) 海岛是什么地方，为什么不能步行前去呢？是否途径是存在的，只是迄今为止不被人发现而已？

(2) 以往他见到过的地方，处处都是可以步行而去的。为什么海岛就不同？遵循过去生活的经验，还会有错吗？

(3) 天下无难事，只怕有心人。他认为自己不但有心，而且肯努力，难道有这么多的优良条件，他仍无法达到目的吗？

只要有一定地理知识，不能步行去海岛是显而易见的。但这位父亲不了解，而且受困扰，这只是因为他缺乏地理知识的缘故。尝试用圆规和直尺去解答三大难题，正像这位父亲尝试步行去海岛一样，都是不可能的。换句话说，根本不存在一个做法使三大难题中的任何一个得到尺规解答。并非“做法是存在的，只是没有被发现而已”，正像步行到海岛的途径根本不存在一样。

正因为数学界相信三大几何问题早已被证明是尺规不可解问题，因此数学家对那些声称自己找到三大几何问题尺规作图方法的人采取置之不理的态度是完全正当的。做出声称的研究者们往往指责说，数学家在未明确指出其错误之处的情况下就简单否定其成果的做法是不虚心、不合理的。但实际情况是，数学家在清楚知道找这些做法的错误之处完全是无聊、繁杂和白白浪费精力的工作时，他们并没有责任为此而浪费自己宝贵的时间。对此，华罗庚先生指出在这桩讼案中，真正需要改正做法与态度，需要更虚心，需要屈尊降贵了解对方研究成果的，不是数学家们，而是业余研究者们。

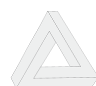
业余研究者们还有一个自认为有力的证据证明自己可能正确：许多重要的数学发现在最初也是被数学界所忽视或拒绝接受的。诚然，在数学发展史上这样的例子是很多的，比如上一章中我们介绍过的阿贝尔与伽罗瓦。但问题是，为数学界忽视的研究并非都有价值，众多业余研究者所奉献的数不清的可笑的“论文”就是毫无价值的。如何鉴别出总会发光的金子与永远都不会发光的垃圾呢？一个重要的参考依据是：研究者是否站在巨人的肩膀上。如我们上一章所介绍到的，阿贝尔与伽罗瓦的发现都是建立在向大师学习的基础之上的，而能极其迅速地汲取数学大师们的成果正是其数学天才的体现之一。可以说，任何伟大的数学家莫不如是，而任何一项伟大的数学发现也都是建立在前人研究基础之上的。至于那些不肯学习前人的结果，却试图通过向壁虚构而奇迹般独步古今中外的研究者，可以断定其研究结果是没有任何价值的。

有意思的是，任何清醒剂都会在部分这类研究者面前失效。对这些昧于现代数学知识背景的数学狂怪而言，任何逻辑都无法说服他们，他们几乎听不进什么道理。当然，这种死不认错、蛮干到底的人注定不会在数学界留下任何建树，而只会成为数学界的笑柄。

本章最后有必要指出，证明问题的“不可解”并非一个令人失望的答案，而认识

到有些事情确实是不可能的，往往标志着数学思想的一大飞跃。可以看到，上一章介绍的“五次及高于五次代数方程不存在根式解”，这一章介绍的“三大几何问题是尺规不可解问题”，都是通过引入新的数学思想与工具后才最终获解的。事实上，这种不可解问题，在数学上带来了极重要且有价值的思想：怎样才能证明某种问题是不可解的？在上一章与本章中，我们已经两次接触到这种新思想。在下一章中，我们将与其再次相逢。





第三章

欧几里得第五公设问题

第一节
第五公设问题的由来

公元前 3 世纪，古希腊著名数学家欧几里得完成了数学史上的光辉巨著《几何原本》。在这本伟大著作中，欧几里得创造性地使用了一种公理化方法，他从一些不加证明的公理、公设出发，构建起一座几何学的大厦。在这些作为起点的命题中，独特的“第五公设”（也称“平行公设”）后来成为“科学史上最著名的一个命题”。本章中我们要介绍的正是人类在 2000 多年时间里对这一公设的漫漫探求之路以及由此导致的几何学革命。

在本节中，我们需要先去了解曾作为数学“圣经”的《几何原本》。正是它的问世，才带来了我们要探讨的曾使无数人着迷并困惑的欧几里得第五公设问题。

数学“圣经”

日出东方。数学起源于古埃及、古巴比伦等东方文明。在漫长的岁月中，播下数学火种的东方人慢慢积累起不少数学知识，使数学获得了初步

发展。不过，他们对数学知识的认识是建立在直觉与经验的基础上的，而数学也主要是一种计算的实用工具。直到约公元前600年，数学的本质才在另一个民族——古希腊人——手中发生了改变。经验的算术和几何方法逐步被加工升华为具有初步逻辑结构的论证数学体系。

其中最关键的一步是，古希腊人对古代流传下来的知识开始追问为什么并尽力回答为什么，从而将证明的思想引入数学。他们的这一努力将数学提升到一种新的境界，从此数学从具体的、实验的阶段过渡到抽象的、理论的阶段，并逐渐形成一门独立的、演绎的学科。这在数学史上是一次不寻常的飞跃。

人们通常认为泰勒斯（约公元前624—前547）是这一新数学思想的开创者，相信是他沿着论证数学的方向迈出了第一步。泰勒斯在当时以其广博的知识与智慧赢得了崇高的声誉，被尊为“古希腊七贤之首”。作为西方历史上第一位数学家，泰勒斯极力主张对几何陈述不能仅凭直觉上的貌似合理就予以接受，而必须要经过严密的逻辑证明。泰勒斯还在命题证明的思想指导下，确立和证明了第一批几何定理，从而使他成为论证数学之父。

在泰勒斯第一个引入命题证明的思想，开创了某种朴素的几何证明之后，古希腊著名哲学家、数学家毕达哥拉斯（以毕达哥拉斯定理闻名）及其学派大大推进了这种思想，促使数学沿着这一新方向成长。他们研究并证明了更多数学命题，包括一些深奥得多的数学结果。可以说，毕达哥拉斯及其学派在使几何学从经验上升到理论方面做出了关键性贡献。古希腊几何也由此真正转向以演绎证明为基础。

期间，古希腊人还由于无理数的引入经历了第一次数学危机（参见本人拙作《数学悖论与三次数学危机》）。经过这次危机的洗礼，希腊人更加确认：直觉、经验乃至实验都不是绝对可靠的，推理论证才是可靠的，证明的思想在希腊人心中扎下了根。进而，古希腊人发展了逻辑思想并加深了对数学抽象性、理想化等本质特征的认识。

在促使这些数学思想牢固确立的过程中，我们在上一章中已经提到的伟大哲学家柏拉图（约公元前 427—前 347）起到了重要的作用。他深受毕达哥拉斯学派影响，并发展了他们的观点。他的数学思想对希腊产生了强有力的影响。柏拉图强调要把数学奠基于逻辑之上，并坚持使用准确的定义、清楚的假设和严格的证明。他坚持对数学知识进行演绎整理，并强调“应从自明的假设出发进行严格的证明”。这一思想后来成为古希腊公理方法的发端。

柏拉图的思想在他的学生亚里士多德那里又得到了发展和完善。作为“古希腊最博学的人”，亚里士多德研究的领域极其广泛。对数学而言，他最大的贡献是在前人基础上完成经典著作《工具论》，把逻辑规律典范化、系统化，阐述了逻辑学理论，从而创立了古典逻辑学。

他研究和讨论了三段论问题，他相信，逻辑论证应该建立在三段论的基础上。所谓三段论，就是“由三个判断构成，其中两个判断是前提（大前提和小前提），一个判断是结论”。举一个简单的例子：所有人都会死（大前提）、柏拉图是人（小前提），柏拉图会死（结论）。如果三段论的前提正确，那么结论也必定正确。在亚里士多德看来，不是任何知识都可以作为三段论的前提，大前提必须是大众普遍接受的事实。他还对每门特殊学科中的基本原理和大众共知的普遍真理加以区分，把前者称为公设，后者称作公理。他举出“两个等量减去同一量后还剩下两个等量”这一公理作为普遍真理的例子。他认为，根据我们不会出错的直觉可知，公理为真。另外，亚里士多德清楚地认识到，我们必须有这些真理作为我们进一步推理的基础。因为任何推理都必须建立在某些前提下，而这一过程不能无穷重复下去。亚里士多德还指出，有些概念必须不给出定义，否则就无起点可言。

亚里士多德的逻辑思想进一步为把几何整理在严密的体系之中奠定了基础，公理化思想已是大势所趋。从公元前 6 世纪起，经过几个世纪的积累后，希腊人也已经获得了大量的数学知识并证明了许多数学命题。一个严整的几何结构已是“山雨欲来风

满楼”了。不久后，亚里士多德的逻辑学与当时已有数学成果完美结合，孕育出了欧几里得的巨著《几何原本》。

欧几里得（约公元前330—前275）是公元前3世纪古希腊著名数学家。关于他的生平，现在知道的很少。他早年大概就学于雅典，精通柏拉图的学说。公元前300年左右，在埃及托勒密王的邀请下，他来到亚历山大，长期在那里工作。我们可以引述关于他的两则轶事来说明他是一个什么样的人。有心的读者会从这两则故事中得到一些重要启示。



欧几里得

一则故事是，托勒密王问欧几里得，除了他的《几何原本》一书外，有没有其他学习几何的捷径。欧几里得回答道：“几何无王者之道。”不肯刻苦钻研、总打算投机取巧的人，可以把这句后来传诵千古的学习箴言作为自己的座右铭。

另一则故事是：一个学生刚开始学习第一个命题，就问欧几里得，学了几何学之后将得到些什么。欧几里得说，给他三个钱币，因为他想在学习中获取实利。由此可

见，欧几里得是反对狭隘的实用观点的。

欧几里得这位伟大的几何建筑师在前人准备的“木石砖瓦”材料的基础上，天才般地按照逻辑系统把几何命题整理起来，建成一座巍峨的几何大厦，完成了数学史上的光辉著作《几何原本》。这本书的问世，标志着欧氏几何学的建立，在数学发展史上树下了一座不朽的丰碑。欧几里得也因此在数学史上获得了显赫的名声，被后人称为“几何学之父”。

使《几何原本》赢得非凡评价的不仅仅是其内容的重要，或者其对定理出色的证明，更为重要的是欧几里得在书中创造的被称为公理化的方法。

在证明几何命题时，一个命题总是从前面的命题推导出来的，而前面的命题又是从更前面的命题推导出来的。我们不能这样无限地推导下去，应有一些命题作为起点。这些作为起点、具有自明性并被承认下来的命题称为公理，如中学所学的“两点确定一条直线”等。同样，对于概念来讲，也有些不加定义的原始概念，如点、线等。在一个数学理论系统中，我们尽可能少地选取原始概念和不加证明的一组公理，以此为出发点，利用纯逻辑推理的法则，把该系统建立成一个演绎系统，这样的方法就是公理化方法。欧几里得正是采用这种史无前例的陈述方法，来组建整个几何学体系的。

书中第 1 卷开始给出 23 个定义，涉及点、线、面、角、圆、三角形、四边形等。我们列举有关点、线、面的定义，可以从中窥其大略：点是没有部分的；线只有长度而没有宽度；一线的两端是点；直线是它上面的点一样地平放着的线；面只有长度和宽度。

接着是 5 个公设。

公设 1：由任意一点到另外任意一点可以画直线。

公设 2：一条有限直线可以继续延长。

公设 3：以任意点为心及任意的距离可以画圆。

公设 4：凡直角都彼此相等。

公设 5：同平面内一条直线和另外两条直线相交，若在某一侧的两个内角的和小于二直角的和，则这二直线经无限延长后在这一侧相交。

随后是 5 个公理：等于同量的量彼此相等；等量加等量，其和仍相等；等量减等量，其差仍相等；彼此能重合的物体是全等的；整体大于部分。

欧几里得对公理、公设的区别是采用了亚里士多德的观点，即公理是适用于一切科学的真理，而公设则仅适用于几何。现在我们一般已不再区分公设与公理，而统称为公理。

《几何原本》后面各卷又给出了许多定义，最终全书给出了 119 个定义，但没有再添加新的公理或公设，因此上面给出的已是全书所需的全部公理和公设了。119 个基本定义、5 条公设和 5 条公理成为全书推理的出发点、论证的依据。将公理、公设、定义作为已知，欧几里得先证明了第一个命题，然后又以之为基础，作为新的已知来证明第二个命题。如此循序渐进，欧几里得有条不紊地由简单到复杂最终推出共 465 个命题（即现在所说的定理），其中包括 54 个作图题。其论证之精彩，逻辑之周密，结构之严谨，令人叹为观止。零散的数学理论被他成功地编织为一个从基本假定到最复杂结论的连续网络。因而在数学发展史上，欧几里得被认为是成功而系统地应用公理化方法的第一人，他的工作被公认为是最早用公理法建立演绎数学体系的典范。

《几何原本》是古希腊数学成果、思想、方法和精神的结晶，它的问世是整个数学发展史上意义极其深远的大事。欧几里得所树立的这一数学史上的理论丰碑深刻地影

响了日后的数学，以公理为前提来获取知识和使用证明来得到新结论的原则成为数学家们的典识和所有数学的规范，对西方数学的发展产生了不可估量的影响。作为整个科学史上发行最广、使用时间最长的书，《几何原本》的手抄本曾统御几何学 1800 年之久，印刷术发明后，又被译成多种文字，共有 2000 多种版本。它的影响还越出数学，对西方思想产生了极为深远的影响，成为整个人类文明史上的里程碑。这一切，使得《几何原本》一书在其问世后的 2000 多年中成为数学中的“圣经”。

欧氏几何的污点？

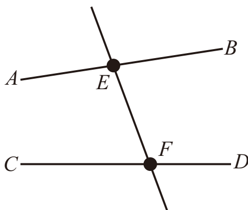
公元前 3 世纪，欧几里得精心雕琢完成了在数学发展史上具有极其深远影响的数学巨著《几何原本》，树立了一座“希腊理智最完美的纪念碑”。在其后很长的时间中，这本巨著被视为“绝对真理”的化身。然而美玉有瑕，这本巨著自问世起就有一个疑问萦绕在人们心头。

通过上面的介绍，我们已看到欧几里得的几何学以 5 条公理与 5 条公设作为逻辑推演的前提，其整个大厦都建筑在它们之上。这些公理与公设中，有 9 条看上去极为显然，极易让人理解并接受。唯独第五公设显得非常特别。让我们再来回顾一下这一公设所陈述的内容。

公设 5：同平面内一条直线和另外两条直线相交，若在某一侧的两个内角的和小于二直角的和，则这二直线经无限延长后在这一侧相交。

通过下图，我们理解一下这一公设表述的内容。已知同平面内一条直线 EF 和另

外两条直线 AB、CD 相交，并且左侧两个内角的和（即 $\angle AEF+\angle CFE$）小于 180°，于是可得出结论，直线 BA 与 DC 在延长后必定会在左侧相交。



显然，与其他公理、公设相比较，这一公设在叙述上冗长、复杂，看上去不是那么一目了然。它似乎缺乏公设、公理应有的简单性和自明性。它的引入使数学家们感到困惑。困惑的产生并非因为有人怀疑这一公设的真理性，而是因为人们普遍认为这个公设是逻辑的必然，是正确的，并且人们都确信这一公设是实际的或物理的直线行为的理想化，它与其他公理、公设一样是对真实物理空间的真实描述。人们真正置疑的是：为什么要把它作为公设。换言之，人们怀疑第五公设不是一个公设而是一个可以证明的命题，因而是一条定理。用一位古希腊数学家普罗克洛斯的话就是："公设 5 完全应从公设中剔除，因为它是一条定理……"

欧几里得本人对此公设的使用加深了人们对此的怀疑。人们注意到，欧几里得在《几何原本》一书中一直尽量避免使用这一公设。他在不依靠第五公设的情况下推出前 28 个命题，直到证明第 29 个命题（一条直线与两条平行直线相交，则所成的内错角相等，同位角相等且同旁内角的和等于二直角）时才第一次用到了这一公设。这使人们做出合理的推测：欧几里得之所以把它列为公设，不是它不可证明，而只是因为他本人没有找到证明。如果是这样，这实在是这部千古不朽巨著的白璧微瑕。因为在一个完美的公理体系中，人们希望能在命题的丰富程度不变的情况下，把公理、公设的数目减至最少，消除所有的多余公理、公设。因此，如果能证明第五公设真是多余的，那么欧几里得几何学的公理、公设数目将会减少，这将洗刷掉欧几里得几何学的"瑕疵""污点"，完善欧几里得几何，并使欧几里得几何学的逻辑结构和美学魅力得到更大提高。

第二节
第五公设的试证之路

能否把第五公设的公设资格取消掉？这个诱人的想法吸引了欧几里得以后无数数学家。在欧几里得时代之后2000多年的漫长岁月中，出于似乎非常充分的理由，数学家们开始寻求公设5的推导。像数学家鞋子里的小石子一样，第五公设问题纠缠着一代又一代的数学家。数不清的人关注它并与其展开过斗争。结果却是，无数的尝试与努力都失败了，第五公设呈现出某种神话特征，并成为科学史上最著名的一个命题。

第五公设的等价命题

在摆脱第五公设（也称平行公设）困扰的努力中，第一个有影响的工作是由古希腊天文学家托勒密完成的。在这次认真的尝试中，托勒密采取的方式是直接证明法。他试图通过欧几里得的其他9个公理、公设直接推导出第五公设。当然，在寻求证明的过程中，可以自由地应用除第五公设以外的公理、公设，还可以自由使用欧几里得不借助第五公设获证的

前28个命题。

托勒密认为自己成功了。但在证明过程中，他假定了：两直线平行后，另一与之相交直线一侧内角成立的东西，也必在另一侧同样成立。后人指出，这一假定事实上等价于第五公设。利用等价于第五公设的命题去证明第五公设，这是一种循环论证。这样的证明被宣告是无效的。

托勒密的错误在试图直接证明第五公设的尝试中极为常见，循环推理愚弄了许许多多试图证明它的人，包括一些著名的数学家。下面这个故事就是一个相关的例子。

1812年冬，法国著名数学家拉格朗日要在法兰西科学院做一个关于欧几里得第五公设的报告。他对那天即将发表的演讲很有把握。然而就在演讲前一刻，拉格朗日对自己的证明产生了疑问，感觉有什么东西错了。这时数学家勒让德已把他介绍给了听众。拉格朗日焦虑地走上讲台。在讲台上，他一言不发，只是徒劳地摸索着他的笔记，这样持续了一段非常长的时间。最后，他抬起头来注视着听众，用惯常的虚弱嗓音和意大利口音，平静地发表了一个声明，请求大家原谅自己。他声称自己的证明存在问题，然后他走下讲台，走出了会堂。

故事中提到的法国著名数学家勒让德（1752—1833）在同样的问题上写下了更典型也更悲壮的篇章。

勒让德在初等数学史上的名声主要源于他那本普及而影响深远的《几何学基础》。在这本书中，他改进了欧几里得《几何原本》的讲授方法，简化并重新编排了许多命题。他的通俗版本备受人们欢迎，并成为后来初等几何教科书的蓝本。在这本书的第一版（1794）中，勒让德给出了自己对平行公设的证明。其后一直到1823年，在大约30年中，勒让德一次又一次地做出徒劳的努力。一本数学史书中对此描述道：

“几乎在他的《几何学基础》每一个版本中，勒让德都‘证明’了平行公设。然而，每一个证明都因为不充分而受到其他数学家的攻击。勒让德却是惊人的固执，他

拒绝考虑‘证明’平行公设的方法有可能是错误的，他总是在后来的版本中提供新的证明，希望他的批评者们能感到满意。在第三版（1800）中，他用三角形内角和来代替原来的证明，在第九版（1812）中，他放弃了这个证明，就像他后来解释的那样，‘又回到了欧氏简单的证法上来，参考了一些严格证明的注释’。他尤其关心的是找到一个适合学生学习的证明。在第十二版（1823）中，他相信自己发现了正确的证明，这个证明保留在他在世时出版的其他所有版本之中（它当然也是错误的）。”

事实上，这样的无效证明反反复复出现。许多数学家在证明过程中，不自觉地运用或假定了某个新的公设，在这个新的前提下推导出了平行公设。然而，这个新公设事实上只是平行公设的等价形式，是伪装起来的平行公设而已。这种努力的最终结果不是证明了第五公设，而只是给出了第五公设的某些等价命题。

另一类尝试途径是，明确定义一个新的公设来取代欧几里得的平行公设，而不像前者那样把它隐藏在一个新定义或证明过程中。许多提出的公理在直观上似乎更加不证自明一些，所以它们的创造者认为他们已经达到目标，然而进一步检查却看出这些代替公理也不能真正令人更满意。虽然在这些替代方案中，有的在陈述上可能比欧几里得第五公设简洁、清晰，但是否更确凿可信却是仁者见仁的事情。在所有这类替代方案中，最著名也是我们学习中经常采用的是：过直线外一点有且只有一条直线与该直线平行。这一受到人们偏爱并为我们所熟悉的表述通常被称为普莱费尔公理。普莱费尔（1748—1819）是当时有一定名气的苏格兰物理学家和数学家。1795 年，他编辑了一版欧几里得《几何原本》，欧几里得原来复杂而冗长的平行公设被改写成这一简洁明了的等价形式。普莱费尔的名字就凭这个表述流传到现在。

在漫长的岁月里，对上述这两类尝试的努力一直持续不断地进行着，但没有人获得真正成功。不过，这类努力也并非全无所获，因为在这一过程中，人们得到了许多在逻辑上与第五公设等价的命题，由此加深了对第五公设的认识。

除普莱费尔公理外，下面再列出一些与平行公设等价的命题。

1. 两条平行线之间距离处处相等（等距公设）。
2. 在已知直线同侧与它等距离的点组成一直线。
3. 一直线的垂线和斜线总相交。
4. 三角形的内角和等于两直角（这一命题还可弱化为：存在一个三角形，其内角和等于两直角）。
5. 相似三角形存在。
6. 一四边形若一对对边相等，且与第三边都成直角，则其他两角都是直角。
7. 一四边形若有三个内角为直角，则第四个内角也是直角。
8. 矩形存在。
9. 勾股定理成立。
10. 三角形的面积不会都小于一个固定值。
11. 任意给定三个不共线的点，存在一个圆通过这三个点。

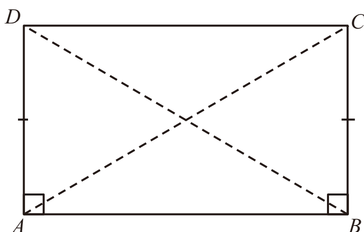
新几何的先行者

由于正面的努力没有成功，到 18 世纪的时候人们开始使用间接证明的方法，即尝试用平行公设的否命题来代替原来的平行公设，并研究由这样的公设（与其他公理公设一起）出发是否会得出矛盾。显然，一旦得出了矛盾，那么依据归谬法，平行公

设就得到了证明。沿这条途径取得的进展表示一门新几何正在孕育中，而在这方面做出贡献的几位数学家也因此成为新几何的先行者。下面我们简单介绍其中几位的工作。

在试图“清晰地证明这条颇存争议的欧几里得公理”的过程中，第一个走上间接证明道路的是意大利数学家萨凯里（1667—1733），他首先应用归谬法来证明这一著名公设。

萨凯里考虑了一种四边形，如图所示，$\angle A=\angle B$ ＝直角，$AD=BC$，这一四边形后被称为萨凯里四边形。



萨凯里在不用第五公设的前提下，容易地推出了$\angle C=\angle D$。于是，这两个等角的大小可能出现三种情况：两者皆为直角，两者皆为钝角，两者皆为锐角。萨凯里把它们分别称为直角假设、钝角假设、锐角假设。

萨凯里清楚直角假设等价于平行公设。因此为了证明平行公设，可以走一条迂回路线：证明钝角假设与锐角假设不成立。如果能成功地排除钝角假设与锐角假设的可能性，那么就可得出两者皆为直角的结论，从而证明平行公设。

为此，萨凯里使用了反证法。他假设钝角假定与锐角假定成立，看能否推出矛盾。

从钝角假定成立出发，在假定直线为无限长的情形下（欧几里得同样含蓄地提出过这样的假定，对此似乎没有人提出异议），萨凯里很容易推出了矛盾，因而钝角假定可以去除。

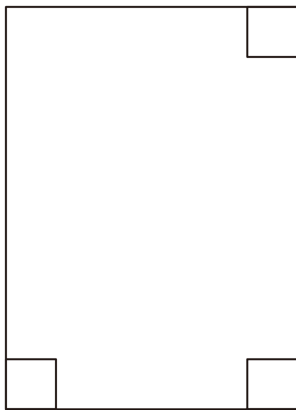
但从锐角假定成立出发，萨凯里导出许多稀奇古怪的命题，比如三角形三内角之和小于两直角等。特别在得到某一结论后，他声称“由于与直线的自然特性相矛盾，锐角假设是完全错误的”。

由此，萨凯里确信自己达到了最初的目的，即他证明了欧几里得第五公设，从而辩明并免除了欧几里得几何中一直存在的这一大污点。于是，去世前几个月，萨凯里把研究结果整理成《欧几里得无懈可击》(1733）一书发表。他相信，在他的工作后欧几里得几何学真正完美无缺了。

确信已经更换了《几何原本》基础的萨凯里相信自己完全可以含笑九泉。然而在他死后 30 年，德国数学家克吕格尔（1739—1812）在其博士论文中指出：萨凯里并没有得出矛盾，而只是得出了奇怪的异于经验的结果而已。同时，克吕格尔还对第五公设能否被证明表示了怀疑。

萨凯里与克吕格尔的研究启发了另一位数学家：兰伯特。

兰伯特（1728—1777）是德国数学家。1766 年，他完成《平行线理论》一书。在书中，兰伯特考虑的是三个内角是直角的四边形。



从这一基础图形出发，兰伯特考虑第四个内角。显然对此可作直角、钝角和锐

角三个不同的假设。他还证明了三个假设分别等价于三角形内角和等于、大于和小于 180°。

从钝角假定成立出发，与萨凯里一样，在假定直线为无限长的情形下，兰伯特推出了矛盾，从而否定了钝角假设。

但在否定锐角假设时，他遇到了很大的困难。他从锐角假设出发，推导出各种结果。但他极力寻求锐角假设下可能出现的逻辑矛盾时，却始终未能达到目的。对得到的这些不包含矛盾的新命题，他没有轻率地加以否定。在努力的最后，他声称“这条假设一点也不自相矛盾”“应当有理由认为，锐角假定远没有遭到反驳”。

在其书中，兰伯特还提出一些重要思想。他证明了，钝角假设下大于两个直角的超出量，与在锐角假设下小于两个直角的亏量均与三角形的面积成正比，并且根据这一结论，他推测钝角假设下的几何学对应着球面几何学，锐角假定在一种虚球面上成立。他的推测后来得到证实。

此外，他还具有一种先进的观点。他意识到任何一组不导致矛盾的假说都提供了一种可能的几何学。这样的几何学将会是逻辑上有效的结构，即便可能与物理形状关系不大。

然而他似乎深信欧几里得的几何学是空间真理，当他感到自己难以成功地反驳锐角假设后，便放弃了关于欧几里得平行公设的研究。并且很可能是因为对自己的结论并不满意，他生前没有出版他的《平行线理论》。直到他去世 10 年后，他的朋友在 1786 年出版了此书。

在兰伯特之后，一位业余研究数学的法学教授施韦卡特（1780—1859）又向前迈进了一步。1807 年，他出版《平行线定理》。1816 年他写了一份备忘录，并于 1818 年送交高斯征求意见。施韦卡特在自己的研究中区分了两种几何：欧几里得几何与假设三角形内角之和不是两直角的几何。他也推导出一些在后一种几何中成立的命

题，其实这些与萨凯里和兰伯特根据锐角假定推导出的结论是一致的。因为觉得这种几何可能在星空中成立，所以施韦卡特把它称为“星空几何”。施韦卡特的外甥陶里努斯（1794—1874）继续研究星空几何，又得到一些新的结果，并发表《平行线理论》（1825）和《几何原理》（1826）。重要的一点是，施韦卡特与陶里努斯都不仅推导出这种新几何中的定理，而且他们与兰伯特一样开始相信并接受：可以选取与平行公理相矛盾的其他公理来建立逻辑上相容的新几何。

第三节
非欧几何的诞生

18 世纪后，我们看到几位非欧几何先驱者几乎已触摸到新几何脉搏的跳动。然而，由于深受传统几何观念的束缚，他们丧失了发现新几何的契机。突破具有 2000 年根基的欧几里得几何传统的束缚，创立新的几何观念，还需要更多巨人出场。奇妙的是，“犹如紫罗兰在春季到处开放”，19 世纪 3 位不同国籍的数学家几乎同时爆发灵感，各自独立地迈出了决定性的一步。

从乌有创造一个新奇的世界：不同凡响的二十几页

“你千万不要去碰第五公设问题。我知道这将带来什么后果。我曾经经历过这一无底的黑暗。它熄灭了我一生的所有光明和乐趣。我恳求你放弃第五公设的研究。我想我已经为真理做出了牺牲。我已为除去几何学的瑕疵并使其更加纯净而奉献出了我的一生。我已做了大量的工作，我的成果远远超出他人。然而我仍没有达到令人满意的

结果……当我看到，没有一个人能达到黑暗的尽头时，我就回头了。我满怀痛苦地回头了，为自己也为人类而怜惜。我似曾到过这些地区，似曾驶过这无底死海的每一块礁石，可是每一次都帆破桅折而归，我曾想也不想就以我的生命和欢乐来冒险，我的毁灭和失败就是这样来的。我回过头来，感到不安，可怜自己也可怜所有的人……你必须像憎恶一切恶行那样憎恶它；它会剥夺你的一切：你的乐趣、你的健康、你的休息，还有你一生的全部幸福。这无底的黑暗深渊将吞噬一千个牛顿；它永远不会在大地发光……”

这是父亲写给儿子的信中的一段。信中自称对欧氏平行公设问题探讨了大半辈子而毫无所获的写信者是匈牙利数学家 F. 波尔约（1775—1856）。

1796 年，21 岁的 F. 波尔约进入哥廷根大学学习，在那里他与著名数学家高斯成为同学与挚友。两人常在一起，交流数学思想或结伴长途旅游。有一个故事说，一次高斯把这位朋友带回自己家。高斯的母亲把 F. 波尔约单独叫到屋外，向他询问，她儿子高斯是否如每个人说的那么聪明，如果确实如此，那么她儿子的前途如何。F. 波尔约回答说，高斯必将成为欧洲最伟大的数学家。一直为儿子感到自豪和高兴的母亲闻言眼泪一下子流了出来。有一句广为流传的对高斯赞美的话也是出自 F. 波尔约之口，他称高斯是“数学的巨人，从高高的苍穹之下，看尽群星和渊薮”。

F. 波尔约在 1804~1853 年一直担任一所学院的数学、物理学和化学教授，但他兴趣很广泛。他曾花大量时间写悲剧。他爱好诗歌，也喜欢音乐，喜欢拉小提琴。他还有一个奇特的业余爱好，就是制造设计新颖的烤炉。在数学方面，他完成过大量基础数学的论文。其数学代表作《写给好学青年的数学原理》试图建立一个坚实而系统的几何基础及算术、代数、分析的基础。在几何基础方面，他特别注重欧氏平行公设的研究，如其信中所说，为此他几乎耗费了一生的精力，然而他所收获的只是失望。然而更令他失望的是他眼看着自己的儿子 J. 波尔约竟然也迷上了这个问题。





F. 波尔约和 J. 波尔约

J. 波尔约（1802—1860）在少年时代就在父亲的辅导下学习了微积分、分析力学等高深学科的基础知识。1818 年，他进入维也纳皇家工程学院学习。1822 年毕业分配至军事部门，从事军事研究工作。与父亲一样，J. 波尔约是一个多才多艺的人。他后来以有才华的数学家闻名，但鲜为人知的是，他还是一个充满激情的小提琴奇才和技艺高超的击剑手。关于他一个流传最广的故事是，J. 波尔约在军营和骑兵军官们一起时，因为某种矛盾，13 个官员同时挑战他。波尔约答应了，不过要求在同一个人决斗之后他要用小提琴拉一小曲。结局是，13 场决斗他都赢了，而输者与他一同共享了 13 首小提琴曲。

约 1820 年，尚在大学的 J. 波尔约就已开始研究平行公设，而他对此的兴趣无疑正来自他的父亲。当得知儿子着迷于第五公设的研究后，早已对此心灰意冷的 F. 波尔约在 1820 年给儿子写了上面那封语重心长的信，要求儿子放弃这种无望的研究。

父亲的警告对儿子没有产生影响。J. 波尔约继续自己的研究。最初，他从正面入手，试图用欧氏其他公设来证明平行公设，结果失败了。在对自己所作研究进行认真反思后，他开始转向用归谬法证明，即从反面来考虑命题，看否定平行公设能否引出与欧氏几何其他公设或公理相悖的结果。在“已知一条直线和该直线外的一点，过这

点存在无数多条与已知直线平行的直线”的假定下，经过严密推理，他推出一系列全新的无矛盾的结论。随后，他提出别开生面的思想：第五公设是一条独立的公设，若以自己的新假定替代“第五公设”，便可以构建一门独立的新几何，从而创造出一个新的世界。

经过几年的艰苦努力后，独辟蹊径的 J. 波尔约于 1823 年完成论文《空间的绝对几何学》，时年 21 岁。论文中，J. 波尔约构造出一种他称为绝对几何的新几何。1823 年 11 月 3 日，J. 波尔约写信给父亲报喜：“我所研究的一定能达到我的目标，只要这一目标是可能的。我虽然还没有成功，但是我发现研究的成果是如此宏伟，令人非常吃惊。对于你的失败，我感到非常同情。亲爱的父亲，当你看到这些结果时，你一定也会这样认为。我现在可以说我已从乌有中创造了一个新奇的世界。”

看到儿子的论文后，F. 波尔约为这种新几何而感到困惑，不能接受。1831 年，在父亲准备出版《写给好学青年的数学原理》一书时，J. 波尔约希望父亲将自己的论文作为该书的附录出版。在儿子的再三请求下，F. 波尔约答应了。1831 年 6 月 20 日，为了听听当时最伟大数学家的意见，F. 波尔约写信给自己的朋友高斯，并随信寄上儿子的论文样稿。但高斯没有回信。1832 年 1 月 16 日他又给高斯去信。1832 年 3 月 6 日，父子俩终于等到了数学之王的复信。

“关于你儿子的工作，当我说我不能称赞他时，你一定会感到震惊……因为称赞他便等于称赞我自己。文章的所有内容，你儿子采取的思路、方法以及所达结果，和我在 30 至 35 年前已开始的一部分工作完全相同，我真是被这些结果吓住了。关于我自己的结果，虽然有一部分已经写好，但我本来是一辈子不愿发表它们的。大多数人对于我们所讨论的问题都抱着不正确的态度，我发现只有少数几个人才感兴趣……我本来打算把它们统统写下来，免得它们和我一同被淹没。让我高兴的是现在可以免去这个劳动了，更让我高兴的是，替代我做这件工作的正是我老朋友的儿子。”

J. 波尔约没有为数学之王信中“头等品质的天才”的赞语而高兴，相反他为高斯信中自称早已发现这种新几何而感到心情沉重，他不相信别人比他更早得到同一结果，他认定高斯在这个发现上要以自己的声望争夺新几何的优先权。为此而沮丧的波尔约陷于失望中。1840 年，另一位俄国数学家写的关于非欧几何的德文著作出版，这使波尔约更为灰心丧气。后来，他基本放弃了数学研究，以致未能在数学上获得更大的成就。

这位在数学上未获当时数学界认可的年轻人在生活中的经历也颇为坎坷。1833 年，他不幸遭车祸致残，于是退役回到父亲家里，不久后母亲去世，而父子在一起并不和睦，两人常有冲突，最后 J. 波尔约迁至偏僻地区过着隐居式的生活。1834 年，他与一当地妇女结婚，生有三个孩子，生活极端艰苦。1856 年，其父去世，同年他又与妻子中断了关系。1860 年 1 月 7 日，J. 波尔约结束了忧伤痛苦的一生。死后葬入无名公墓，此公墓的记事本上登记说：“此人一生没有什么意义。”然而，历史没有遗忘这位非欧几何的发现者。1894 年，匈牙利数学物理学会为他主持修复了墓地，并建造雕像，供人景仰。1905 年，匈牙利科学院高度表彰了他的功绩，颁发了以他的名字命名的国际数学奖，以奖励那些为数学进展做出巨大贡献的人。而他用拉丁文写的，并以附录形式发表的仅 26 页的论文《绝对空间的科学》（1832）亦被称为“思想史上最不同凡响的二十几页”。因“附录”的拉丁文是“Appendix”，后来人们把这篇论文称作“阿乒的克斯”。J. 波尔约终于因自己发现的“新奇的新世界”而永久铭刻在数学史上。

正如 F. 波尔约生动比喻的“紫罗兰在春季到处开放”，在 J. 波尔约之前与之后都有人发现了第五公设的不可证明性和非欧几何的原理，从而共享非欧几何创建者的荣誉。

高斯与非欧几何

在 J. 波尔约发现一个新奇的新世界之前，认识到可能存在一种新几何学的是数学之王高斯。

正如在给波尔约父子的回信中所说，高斯早在 15 岁时就认识到能够存在一种逻辑几何，其中欧氏平行公设不成立。其后，他对这种新几何的认识逐渐深入。1799 年 12 月 17 日，他写信给老朋友 F. 波尔约："我在工作中已有所进展。然而我所选择的路径根本不会引向我们所追求的目标，而你对我言之凿凿已到达这个目标。但我的工作似乎促使我怀疑几何学本身的真理性。的确，我已偶然得到了多数人会当成是（从其他的公理推导出欧几里得平行公设的）证明的许多东西，但在我看来这什么也没证明。"自 1813 年以后，他进一步发展了他的新几何学，最初称之为反欧几里得几何，后又叫超感觉世界几何学，最后改称为"非欧几里得几何"。因此"非欧几何"这个名称正是来自高斯。他深信该几何在逻辑上是相容的。在 1824 年 11 月 8 日写给朋友的一封信中，他写道："假定三角形内角之和小于 180° 会导出一种奇怪的几何，它与我们的欧式几何迥然不同，但完全相容，我已经将其发展得完全满意了。这种几何的定理看起来是悖论，对于没入其道的人来说是荒谬的，然而静心沉思，发现它们绝没有包含不可能的东西。"

但行事谨慎的高斯意识到这种新思想会冲击欧几里得几何的权威，也与人们的常识相悖，还会挑战当时盛行于欧洲的康德哲学，担心这必将遭受世人的攻击和耻笑。在写给朋友的信件中，他坦诚了自己的担心："假如我不保守秘密的话，黄蜂会围绕我的耳朵飞。""我惧怕在我大声讲出我的观点之后，会引起维奥蒂亚人的鼓噪。"维奥蒂

亚人是古希腊一个以愚昧著称的部落，高斯借此影射那些缺乏想象力又不开化的愚钝的会反对他的人。于是，数学之王决定对自己关于非欧几何的发现秘而不宣。在他生前，他关于非欧几何的工作主要通过私人信件透露给朋友，另外就是发表在《哥廷根学报》1816 号和 1822 号上的两篇短评，以及 1831 年与之相关的笔记。1855 年高斯去世后，他的私人信件及笔记公之于世，他关于非欧几何的研究才为人所知，并引起人们的重视。

对公开发表自己的成果顾虑重重的高斯而言，这并非特例。这位立论极端谨慎的数学之王遵循的原则是“少些，但要成熟”“不留下进一步要做的事情”“极度严格的要求”。这些原则使得他的众多思想都没有及时得以发表。1898 年，人们发现了一本高斯的日记。这本著名的数学日记开始于高斯发现正多边形做法的那一天，即 1796 年 3 月 30 日，日记上最后的日子是 1814 年 7 月 9 日。这本日记用密码的形式记录下了许多伟大的数学成果，包括 146 条短条目，除了两条外，其余的都被解释了。对此美国数学史家贝尔评论说：“在高斯死后，人们才知道他早就预见了一些 19 世纪的数学，而且在 1800 年之前已经期待它们的出现。如果他能把所知道的一些东西泄漏，很可能现代数学比目前还要先进半个世纪或更多的时间。”

几何学的哥伦布

在非欧几何的创立中，发现稍晚于 J. 波尔约，但论述最完整深入，并坚决捍卫这种新几何的是罗巴切夫斯基（1792—1856）。

罗巴切夫斯基出生于一个贫穷的家庭。1807 年春，14 岁的罗巴切夫斯基升入喀山大学，从此与这所学校结下了不解之缘。从学生到教授，从系主任到校长，他在这里度过了 40 个春秋。作为一名学生，他听过许多著名教授的课，并以杰出的数学才能赢得了教授们的欣赏。作为一名能干的教师，他教授过许多门数学和物理课程，并于 1822 年成为常任教授。作为一名行政人员，他在喀山大学担任过许多种职位。他曾担任过学校的图书馆馆长和博物馆馆长。在没有任何助手的情况下，他一个人完成了所有工作，使得图书馆和博物馆恢复了秩序。还有一次，俄罗斯政府决定整修喀山大学校园的古老建筑，并建一些新楼。罗巴切夫斯基负责工程质量和经费使用。为了更好地完成任务，他学习了建筑，熟练地掌握了这门学问，结果新建和改建的楼都非常漂亮而实用，而且建设费用比预算的还少。由于他工作卓著，1827 年，深受同事们敬重的他被选举担任喀山大学校长。深爱自己大学的罗巴切夫斯基在成为校长之后，仍会毫不犹豫地做那些最琐碎的杂务。有一则与之有关的有名轶事，一次有个著名的外宾来访，正遇上衣冠不整的校长在擦地板，就把他错当成看门人，请他领着参观学校的图书馆和博物馆。来客对看门人的礼貌和学识大为欣赏，参观结束后，要给他一笔小费。令他惊讶的是，看门人拒绝，来客把这一行为看作一个高傲的看门人的怪癖。当天晚上在领导席上，两人经过介绍相互认识了，于是都向对方表示歉意。由于他的出色工作，数年之后喀山大学成为俄国的一流大学。



罗巴切夫斯基

而作为数学家的罗巴切夫斯基走过的道路则坎坷得多。

罗巴切夫斯基从1815年开始着手研究平行公设。最初他也是循着前人的思路，试图给出第五公设的证明。很快他便意识到自己的证明是错误的。前人和自己的失败从反面启迪了他，于是他调转思路，在“过平面上直线外一点，至少可引两条直线与已知直线平行”的假定下，依照严格的逻辑推理进行推导。结果，他得到一连串古怪却没有任何逻辑矛盾的命题。对此，罗巴切夫斯基大胆提出：“由平行公理否命题出发而得到的结果代表着一种新的几何学，尽管这种新几何学中的许多结论是令人惊异，甚至不可思议的。例如，在这种几何里，三角形的内角之和小于两直角，但是它本身是无矛盾的，它可以同欧氏几何一样成立。”罗巴切夫斯基把自己发现的这种新几何称为“虚几何学”。

1826年2月23日，罗巴切夫斯基于喀山大学物理数学系学术会议上宣读了他的第一篇关于非欧几何的论文《几何学原理及平行线定理严格证明的摘要》。这标志着非欧几何的诞生。然而，他的论文并未引起任何人的兴趣，后来连原稿也被遗失了。1829年，他在喀山大学《喀山通讯》上发表了最早的非欧几何文献“论几何学原理”。其后，他继续自己的研究，并陆续发表多篇论述非欧几何原理及应用的文章。1840年，他用德文发表了《平行线理论的几何研究》。1855年，在双目失明的情况下，他口授完成他最后的著作《论几何学》。然而，罗巴切夫斯基的不懈努力在当时并未获得认可。在创立和发展非欧几何的艰难历程中，几乎一直是孤军奋战的罗巴切夫斯基赢得的不是掌声与声誉，而是如高斯所预料到的“维奥蒂亚人的鼓噪”。

1832年，当他将阐述新几何思想的论文呈送彼得堡科学院审评时，作为审评者的著名数学家奥斯特罗格拉茨基以嘲弄的口吻写道：“看来，作者旨在写出一部使人不能理解的著作。他达到了自己的目的……由此我得出结论，罗巴切夫斯基校长的这部著作谬误连篇，因而不值得科学院的注意。”除了权威的讥讽还有匿名者的攻击。有人以匿名形式撰文对罗巴切夫斯基进行嘲笑与人身攻击：“罗巴切夫斯基先生这位数学常

任教授为了什么重要目的而写一本甚至不能给最差学校的小学教员带来荣誉的书呢?一个教员即使没有学问，至少还应当具有理智，而在新的几何里往往连这一点也很缺乏……”在文章的结尾处，作者讥讽道:“为什么不把标题《几何学原理》写成《对几何学的讽刺》《几何学漫画》呢?”

在逆境中奋斗不止的罗巴切夫斯基始终没能遇到他的公开支持者。当他的德文著作《平行线理论的几何研究》出版后，高斯阅读了这本名著，并私下称赞罗巴切夫斯基是“俄国最卓越的数学家之一”，并下决心学习俄语，以便直接阅读罗巴切夫斯基的全部非欧几何著作。但另一方面，高斯却又不愿公开出面支持罗巴切夫斯基。后来，他积极推选罗巴切夫斯基为哥廷根皇家科学院通讯院士，可是在评选会上和他亲笔写给罗巴切夫斯基的推选通知书中，对罗巴切夫斯基在数学上的最卓越贡献——创立非欧几何却避而不谈。

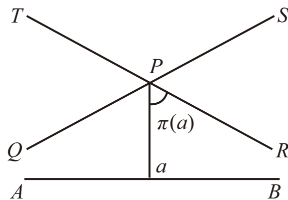
面对各种攻击与挫折，罗巴切夫斯基表现出非凡的勇气。他对新几何的远大前途抱有坚定信念，并在孤境中为其生存和发展奋斗了 30 多年。然而，他的创造性工作在其生前始终没能得到学术界的重视和承认。但历史是公正的。当长期无人问津的非欧几何获得学术界的普遍注意和深入研究后，罗巴切夫斯基的独创性研究开始得到学术界的高度评价和一致赞美，他本人则被赞誉为“几何学中的哥白尼”。罗巴切夫斯基与哥白尼之间确实有着有趣的相似性：都是斯拉夫人，在科学思想和科学观点方面都引起了革命，都是人类宇宙观革命的领袖人物。

如上面所介绍的，J. 波尔约、高斯、罗巴切夫斯基三人彼此独立地发现了非欧几何，并且单就发现新几何的时间上来说，高斯在先，J. 波尔约也略早于罗巴切夫斯基。但是，就论文发表的时间之早，论证之完整，内容之丰富，以及终生不渝宣传、捍卫新几何的功绩而言，高斯与 J. 波尔约都无法同罗巴切夫斯基相比。因此，这一不同于欧氏几何的新几何又被称为罗氏几何。由于三人的基本思想非常相似，下面我们就以罗巴切夫斯基的研究为例，简单介绍一下他们的发现。

罗氏几何简介

罗氏几何与欧氏几何最基本的不同之处在于平行公设。在欧氏几何中，平行公设有多种等价形式，其中最为我们熟知的是普莱费尔公设：过直线外一点有且只有一条直线与该直线平行。在罗氏几何中，平行公设采用了普莱费尔公设的一种否定形式：过直线外一点，至少可以作两条直线和这条直线不相交。这一命题被称为罗巴切夫斯基公理。以此替代欧氏几何中的平行公设，而其他公设都保持不变，则构成了罗氏几何的公理系统。显然，凡是不涉及平行公设的几何命题，在欧式几何中如果是正确的，在罗式几何中也同样是正确的。但由于平行公设不同，在欧氏几何中，凡涉及平行公设的命题，在罗氏几何中都不再成立。事实上，罗巴切夫斯基经过演绎推理引出了一连串与欧式几何内容不同的新的几何命题。

首先，由罗氏平行公设出发可以直接得到下列结果：过直线外一点在同一平面内可以作无穷条直线与给定直线不相交。



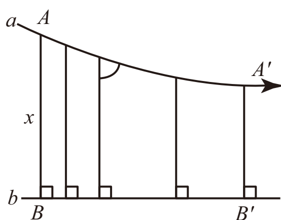
如图所示，根据罗巴切夫斯基公理，过直线 AB 外一点 P，至少有两条直线与直线 AB 不相交。于是，可以把过 P 的所有直线分为两类：一类与 AB 相交，另一类与 AB 不相交。然后可以考虑所有过 P 且不与 AB 相交的直线的极限情形，罗巴切夫斯基证明这样的直线有两条，图中我们用 QS、TR 表示。这两条直线具有特征：它们都不与

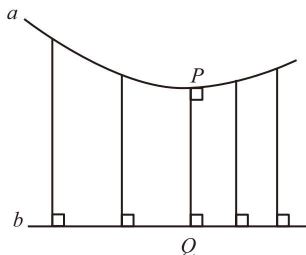
AB 相交，而且它们所成夹角（图中 $\angle SPR$）内的所有直线也都不与 AB 相交，但所有含于 $\angle QPR$ 内的直线都与 AB 相交。也就是说，QS、TR 构成了所有不与 AB 相交的直线的边界。罗巴切夫斯基把这两条线称为 AB 的平行线，而把夹在两线之间与 AB 不相交的直线称为不相交直线（也称发散线）。如果按不相交即平行的通常意义理解，那么我们可以说，在罗氏几何中过直线外一点可以引无穷多条直线与给定的直线平行。

罗巴切夫斯基还把 $\angle QPR$ 的一半称为“平行角”，并证明了这个角是点 P 到直线 AB 的距离 a 的函数，因此可把平行角记为 $\pi(a)$。显然，在欧氏几何中，这个角永远是 $90°$。但在罗氏几何中，这个角随着 a 的变化而变化。具体而言，随着 a 趋于无穷，平行角逐渐减少且趋于 0；随着 a 趋于 0，平行角逐渐增加且趋于 $90°$。

以上述讨论为出发点，罗巴切夫斯基得出了更多不同于欧氏几何的令人惊奇的结果。我们举其中几个例子。

- ❑ 在欧氏几何中，平行线间的距离是常数，但在罗氏几何中情形却大不相同。平行线间不再等距，而且情况还比较复杂。如图所示，若直线 a 与直线 b 是罗巴切夫斯基意义上的平行线，则两直线在平行角的一侧无限地接近，即两者间的距离逐渐变小并趋近于 0，而在另一侧无限地远离，即两者间的距离逐渐增大并趋于无穷。但若直线 a 与直线 b 是罗巴切夫斯基意义上的发散线，则两直线之间的距离在 P 点处最短，此时的垂线 PQ 称为这对分散直线的公垂线，在公垂线两侧这对分散直线都无限远离，即两者间的距离都逐渐增大并趋于无穷。





- 同一直线的垂线和斜线不一定相交。
- 如果两条直线与第三条直线相交且内错角相等，那么这两条直线是发散的。
- 两条平行线与第三条直线相交，在平行方向上的同旁内角之和小于两直角。
- 三角形三条高线不一定相交于一点。
- 三角形内角之和小于 180°，而且随着三角形面积的增大而减小。三角形面积越小，其内角和越接近于 180°；当三角形面积无限增大时，其内角和趋于零。
- 若两三角形的三内角对应相等，则两三角形必全等，即不存在相似而不全等的三角形。
- 不存在面积任意大的三角形。
- 不存在矩形。
- 到一条直线等距离的点的轨迹不是直线，而是一条特殊曲线，叫作等距线。
- 半径无限增大的圆周的极限不是直线，而是一种特殊曲线，叫作极限圆。
- 通过三点并不总能作一个圆，能够作的或者是一个圆，或者是一个极限圆，或者是一条等距线。
- 勾股定理不成立。
- 圆周长与半径不成正比，而是更迅速地增长。

虽然罗氏几何与欧氏几何有如此大的差别，但它却具有一个重要性质：在充分小的区域内，它与欧氏几何差异很小。换一种说法，欧氏几何可看作是罗氏几何的极限情形。因而，如果在罗氏几何中添加上这个极限情形，则它也包括了欧氏几何。在这个意义下，欧氏几何成了罗氏几何的一个特例，而罗氏几何就是一种更普遍的几何学。正因为这个缘故，罗巴切夫斯基把自己的几何体系命名为"泛几何学"，即普遍的几何学。

第四节
非欧几何的发展与确认

罗巴切夫斯基、高斯、J. 波尔约各自独立地提出了本质上相同的一种非欧几何，宣布了非欧几何的诞生。在罗氏几何问世后，新的问题马上摆在数学家面前：除罗氏几何外，是否还存在其他可能的非欧几何，非欧几何真的不存在逻辑矛盾吗？随着非欧几何的发展与非欧几何模型的建立，问题迎刃而解，非欧几何也终于在其创立者去世后获得了数学界的广泛认同。

黎曼几何：非欧几何的发展

1854 年春天，一位名叫黎曼的年轻人为他的前途和即将面临的考试感到忧虑不安。他已经 28 岁了，尚未能自立，他靠父亲每月寄的不多的钱过着贫困的生活。他谦虚地写信给家人说，柏林和哥廷根的最著名的教授们都待他好到不可言状。他已经获得博士学位。现在为获得一个（无薪）讲师聘约，他必须在哥廷根哲学系全体教师面前做一个满意的就职演讲。高斯是黎曼的导师。黎曼已向他提交了三个备选题目。按照常规，导师一般选

择备选题目中的第一个或者第二个。然而，高斯对第三个题目“几何基础”更感兴趣。因为这个深刻而新颖的题目，高斯本人已经仔细考虑了几十年。大约出于好奇，他想知道这位年轻人对这一深奥的问题会讲些什么。“头两个我已经准备得很好了，”黎曼给哥哥的信中写道，“但是高斯挑选了第三个，因此现在我很惶恐不安。”

处于绝境的黎曼，在经过大约七周的时间后，完成了自己的论文《关于构成几何基础的假设》，并在 1854 年 6 月 10 日进行了就职演讲。黎曼的这次历史性演讲很少涉及数学细节，但蕴涵着大量有益于几何学如何发展的真知灼见。演讲结束后，为黎曼论文所震惊的高斯在步行回家的路上怀着不寻常的热情向他的同事韦伯表示了他对黎曼所提思想的最高度的赞赏。历史也已证明，这篇论文确是数学史上的伟大杰作。下面我们先提一下他在这一论文中所建立的一门不同于欧氏几何也不同于罗氏几何的新几何：黎氏几何。

在黎氏几何中，平行公设可表述为：经过给定直线外一点，不存在该直线的平行线。换言之，平面上任何两条直线都相交。事实上，黎曼的平行公设等价于萨凯里等的钝角假设。但在黎曼之前，这一假设被认为与直线可以无限延长（欧氏公设 2）的假定矛盾，因而从萨凯里到罗巴切夫斯基都否定了这一假设。

为了使自己的平行公设成立，黎曼对欧几里得的第二公设做了批判性的审查，对“无限性”与“无界性”这两个概念做了区分。他认为直线可以无限延长并不意味着其长短是无限的，只不过是说，它是无端的或无界的。在这种区分后，黎曼用“直线的无界性”代替了“直线的无限性”。由此，他可以用修正后的公设 2（直线是无界的但长度有限）、修正后的公设 5（平面上任何两条直线都相交），以及欧氏几何的其他公设为基础无矛盾地演绎出第二种非欧几何：黎氏几何。

于是，在通常的欧氏几何之外，我们又添加了两种非欧几何：罗氏几何与黎氏几何，它们统称为非欧几何。为避免误解，这里有必要说明两点。一点是，非欧几何这一术语有几种不同的含义。最狭义的理解是，非欧几何单指罗氏几何；通常意义上，

非欧几何指罗氏几何和黎氏几何；还有一种广义的理解，泛指一切和欧氏几何不同的几何。另一点是，黎氏几何这一术语也应用于两种场合。狭义的理解即是上面我们所指的意思。广义的理解则是后面我们要提到的涵盖面广泛得多的黎曼几何。一种可以避免歧义的方式是采用新的命名方式。1871 年，著名数学家克莱因把欧氏、罗氏、黎氏这 3 种几何进行了重新命名：欧氏几何称为抛物几何，罗氏几何称为双曲几何，黎氏几何称为椭圆几何。如此命名的缘由在于："抛物"一词在希腊文的意思是"既非不足也非过量"，这里它对应于欧氏几何中只存在一条通过已知点与已知线平行的直线的事实；"双曲"一词在希腊文中意指"过量"，这里指的是通过已知点与已知直线平行的平行线的数目，而我们知道在罗氏几何中存在不止一条通过已知点与已知线平行的直线；"椭圆"一词在希腊文中意指"不足"，这里它对应于黎氏几何中不存在通过已知点与已知线平行的直线的事实，用以说明这种几何里平行线太少。

容易明白，三种几何在不涉及平行公设时得出的结论是相同的。事实上，人们把排除平行公设在外的几何称为"绝对几何"。凡属"绝对几何"中的命题，在三种不同的几何中都成立。相反，超出"绝对几何"这一公共部分外的命题，在三种几何学中就截然不同。平行公设成了三种几何的分水岭。平行公设的不同带来了欧氏几何与非欧几何的一些本质上不同的结论。下面我们列举几个。

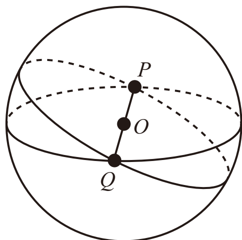
- 三角形的内角和：抛物几何中这一结果总是 180°，它不随三角形大小的改变而改变；双曲几何中这一结果总是小于 180°，而且其大小与三角形的大小有关，三角形的面积越大，内角和越小；椭圆几何中这一结果总是大于 180°，其大小也与三角形的大小有关，三角形的面积越大，则内角和也越大。

- 勾股定理：抛物几何中 $a^2+b^2=c^2$，双曲几何中 $a^2+b^2<c^2$，椭圆几何中 $a^2+b^2>c^2$。

- 圆周与半径的关系：抛物几何中 $c=2\pi r$，双曲几何中 $c>2\pi r$，椭圆几何中 $c<2\pi r$。

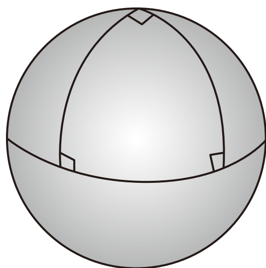
需要指出的是，在充分小的区域内，三种几何的差异是非常小的；区域越小，这种差异也越小。前面我们已经简单介绍了双曲几何，下面借助一个我们非常熟悉的模型——球面模型——来简单理解一下椭圆几何。

我们假定地球是一个完美的球面，并设想其上的居民是只能沿两个方向移动的“球面人”。对他们来说，世界是二维的，只有长和宽，没有高。首先，我们要考虑的是，对他们来说，“直线”（即球面上两点间的最短路线）是什么呢？不可能是我们所熟悉的欧氏几何中的直线，因为这样的线根本不在球面上，而是会像隧道一样穿越地球。事实上，球面上的“直线”应是大圆，其实例有经度圈和赤道。



显然，所有大圆都是相交的，因此球面模型中的任意两条“直线”都相交，这正符合椭圆几何中的平行公设。另外，设想“球面人”决定动身去考察他们的世界。他们沿一条在他们看来是直线（实际是大圆的一段弧）的路走下去，航行刚开始时他们会离家越来越远，但最终他们将会回到他们的出发地。很显然，在球面上适用的不是欧几里得第二公设而是修正后的第二公设：直线无界但有限。另外，容易验证，在球面上，第三与第四公设适用。然而，在宣布球面是椭圆几何的适当模型之前，还有一件事情需要考虑。欧几里得第一公设是说，经过两个不同点有且仅有一条直线。但是，两个大圆总是在一对对蹠点（球面上相对的两个点，如上图中的 P、Q）处相交。因此，经过每一对对蹠点有不止一条“直线”，事实上有无数多条。数学家把这种椭圆几何称为二重椭圆几何。在这种椭圆几何中，两个点并不总是确定唯一的直线。另一种椭圆几何称为单重椭圆几何。在这种几何中，两个点永远确定唯一的一条直线。

在做过以上考察后，我们可以说，球面是（二重）椭圆几何的一个模型。通过这个模型，我们还可以直观地看到在椭圆几何中成立的一些奇怪结论，比如下图三角形的内角和为 270°，大于 180°。



然而从另一方面来说，椭圆几何中的公式一点也不奇怪。事实上，有一个可追溯到古希腊的称为球面三角学的数学分支，它专门研究球面上的三角形和其他形状的性质。而借助于球面三角学，我们可以从普通立体几何学的性质推导出球面的所有性质。换句话说，二维椭圆几何学的公理或定理都可变成三维欧几里得几何学中的定理。

于是，我们看到一件真正奇怪的事情：黎曼的非欧几何就是普通球面上的几何，然而一直就生活在这样一个球面上，并早已发展了球面性质的人类，千百年来却坚定地信奉着一个理想的平直几何空间——欧几里得几何，直到 2000 多年后才由黎曼揭示出球面的非欧几何性！之所以如此的根由在于，由欧几里得到黎曼展现了一种基本观念的跨越。因为我们是三维生物，从古代起我们就能根据普通的三维欧几里得几何学解释二维球面几何。正因为把球面看成是三维空间中的一张曲面，所以古希腊人发展出的球面几何是欧氏的。然而，黎曼所探讨的是没有三维特权的二维“球面人”在面对球面时，如何解释自己生活世界的几何。即黎曼要问的是，“球面人”能否利用数学断定自己生活在平坦的地球上还是在弯曲的球面上呢？黎曼的这一思考方向来自高斯的一个全新思想。在 1828 年发表的《关于曲面的一般研究》中，高斯改变了以前数学家研究曲面总是将其与外围空间相联系的做法，他指出：一个曲面本身就是一个空间，它的许多性质并不依赖于背景空间。为了反映曲面的一般弯曲程度，高斯还引入了曲

面的总曲率（现称高斯曲率），并得到“高斯极妙定理”，即曲面的总曲率是内在的。这意味着，二维地球人如果懂得足够多的数学，他们完全可以在看不到三维空间的情况下，通过测量计算曲率认识到自己居住的球面是弯曲的。高斯的新思想开创了微分几何发展的新时代。这种以研究曲面内在性质为主的微分几何称为“内蕴微分几何”。

黎曼在高斯新观念的基础上，又向前迈进了一大步。他把高斯关于欧氏空间中曲面的内蕴几何推广到任意空间的内蕴几何，并从根本上推广了几何空间的意义。他提出n维流形的概念，引进流形曲率的概念，定义了流形上无限邻近两点的距离（后称为黎曼度量）。以黎曼度量为基础，黎曼建立了一种更为广泛的新几何，现称为黎曼几何。在这一新的几何分支中，选什么来作两点之间的距离即可决定所得之几何。由此，数学空间的种类被成千倍地打开了。理论上而言，无数多种几何学亦可由此建立起来。而其中，比较简单与重要的是曲率处处相同的曲面，即常曲率曲面（空间）。在三维空间中，常曲率空间有三种情形：曲率为正常数，曲率为负常数，曲率恒为零。黎曼指出，这三种情形分别对应于三种几何：椭圆几何、双曲几何、抛物几何。因此，我们说罗巴切夫斯基几何以及欧几里得几何都只不过是黎曼几何的特例。由此，我们也可以说黎曼是最先理解非欧几何全部意义的数学家。他创立的黎曼几何不仅是对已经出现的非欧几何的承认，而且显示了创造其他非欧几何的可能性。

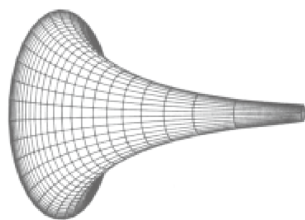
然而，当黎曼进一步发展了非欧几何后，还有一个基本的问题有待解决：椭圆几何、双曲几何这两种非欧几何是相容的吗？虽说暂时没有发现矛盾，但这并不能保证以后人们不会发现矛盾。而一旦有人在这些几何中发现了矛盾，那么它们将变得毫无价值。因此，为了使这两种非欧几何获得认可，还需要证明它们内部不存在矛盾，即是相容的。

事实上，当我们介绍球面模型时，在证明相容性上我们已经迈出了一步。通过建立二重椭圆几何的球面模型，球面几何成为二重椭圆几何的具体表现。或者说，二重椭圆几何中的公理与定理完全可以对应球面几何中的定理。这意味着，如果二重椭圆

几何有矛盾，那么属于欧氏几何一部分的球面几何自然也有矛盾。反之，如果欧氏几何没有矛盾，那么二重椭圆几何也没有矛盾。于是，借助欧氏几何中的球面模型，数学家把二重椭圆几何的相容性归结为欧氏几何的相容性。此外，在球面模型建立不久后，数学家 F. 克莱因提出了单重椭圆几何的一个曲面模型，这又解决了单重椭圆几何的相容性。那么，能否用同样的方法证明双曲几何的相容性呢？为此需要做的是，建立双曲几何模型。

双曲几何模型

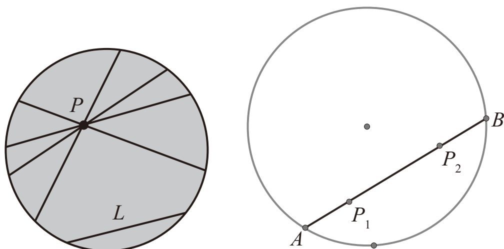
1868 年，意大利数学家贝尔特拉米（1835—1899）在著名论文《非欧几何解释的尝试》中，给出了双曲几何（即罗氏几何）的第一种直观解释。他考虑的是由曳物线绕一条固定轴旋转而生成的旋转曲面——曳物面。这种曲面与球面有一共有性质：在任何一处都有相同的曲率。不过，对球面来说，这个常曲率是正的；而对这种曲面而言，这个常曲率是负的。因此，曳物面常被称为伪球面。



在伪球面上罗氏平行公理成立，即任给定一条直线，过线外一点都有无限多条直线与其平行。当然，这里的“直线”要理解为曲面上两点之间的最短距离。事实上，

在引入适当的“解释规则”后，贝尔特拉米在罗氏几何的平面与伪球面之间建立了对应关系。他还证明了在这样的对应下，罗氏（平面）几何中的所有定理都对应于欧氏（立体）几何中的定理，这样罗氏几何对于欧氏几何的相对相容性就得到了证明。因为如果罗氏几何中包含有矛盾，那么通过这种对应，在欧氏几何中也就可以推出矛盾。由此，伪球面可看作是罗氏几何的一种几何模型。不过，贝尔特拉米实现的并非整个罗氏几何，而是其片段上的几何。因此，他没有解决全部罗氏几何的无矛盾性问题。

这一问题不久就被德国数学家 F. 克莱因解决了。1871 年，克莱因建立起整个罗氏平面的几何模型。其做法是，在普通欧氏平面上取一个圆，并只考虑圆的内部。约定圆的内部为“罗氏平面”，罗氏平面上的点是圆内的点，不包含两端点的圆的弦为罗氏平面上的“直线”。于是，经过直线 L 外一点 P，可以引无数条直线与 L 平行，因为这些线中没有一条在圆内与 L 相交，而我们的约定是只考虑圆内部。因此，在克莱因模型中，罗氏几何的平行公设适用。



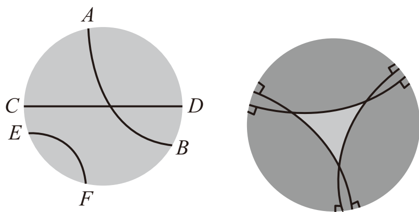
为了进一步将罗氏几何翻译成欧氏几何语言，还必须给出适当的两点间距离和角的大小的定义。克莱因给出圆内两点 P_1 和 P_2 的距离定义如下：$d(P_1,P_2)=\ln\dfrac{P_1B\cdot P_2A}{P_1A\cdot P_2B}$，其中 A、B 是 P_1 和 P_2 确定的直线与圆周的交点。显然，这种距离与我们通常熟知的欧几里得距离大不一样，但是可以验证，它完全具备与距离概念相关的所有性质。另外，在这种定义下，我们可以注意到，当圆内点逼近边界（即圆周）时，距离会变得无穷大。事实上，当 P_1 逼近 A 或 P_2 逼近 B 时，借助对数的性质可以证明，$d(P_1,P_2)$ 变成无穷大。这表明，在这种距离定义下，罗氏直线的长度是无限的。于是欧氏几何中的公

设 2（直线可以无限延长）相应地具有了意义。

经过这样的约定与定义后，可以证明圆内部的点、弦以及其他图形满足罗氏几何的所有公理。于是，圆内部的普通（即欧氏）几何事实就变成罗氏几何中的定理，反过来，罗氏几何中的每个公理或定理也都可以解释成圆内部的普通几何事实。这样，通过克莱因模型，就把罗氏几何的相容性归结为欧氏几何的相容性。

1882 年，法国著名数学家庞加莱（1854—1912）给出另一种模型。在庞加莱模型中，仍取圆的内部（不含圆周上的点）作为罗氏平面；直线有两种，一种是通过圆心的直线（如下图中的 CD），另一种是与圆周正交的任意圆弧（如下图中的 AB、EF）。然后，庞加莱适当定义了两点间的距离与夹角。如同克莱因所做的那样，庞加莱定义的距离可满足通常距离的性质，而且可以使任何圆内点到边界的距离成为无穷大。而两条罗氏直线的夹角则定义为两相应圆弧的夹角，具体而言就是指交点处两圆弧切线的夹角。若罗氏直线是直线段，则切线就是它本身。在经过这样的约定与定义后，可以证明罗氏几何中的公理和定理都可以解释成圆内部的普通几何事实，或者说都是欧氏几何中的特殊定理。这样，通过庞加莱模型，就把罗氏几何的相容性归结为欧氏几何的相容性。

借助庞加莱模型，双曲几何的许多性质可以一目了然。如下面的右图所示，可注意到在圆内部的一个几何结果：三角形内角和小于 180°，而这正是罗氏几何中的一个命题。

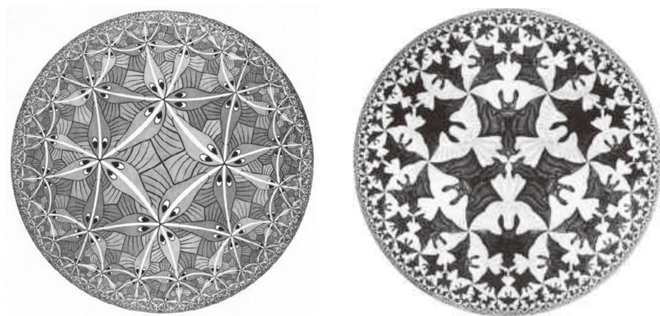


庞加莱模型后来以一种意想不到的方式变得很有名。荷兰艺术家埃舍尔（1898—

1972）被一幅双曲几何的数学图吸引并深感震撼后，被这种几何不寻常的性质所激励，进入到双曲几何的冒险中。在灵感迸发时，他完成了四幅《圆的极限》。通过这几幅图，他成功地将庞加莱双曲几何模型形象化，从而精妙地展现了双曲几何。下面的左图是其中最成功的一幅:《圆的极限Ⅲ》(1959)。埃舍尔本人对这幅作品曾做过精彩描述。

在彩色木版画《圆的极限Ⅲ》中,《圆的极限Ⅰ》中的缺点大部分被克服。我们现在只有“直通”系列，而且属于一个系列的所有鱼都有相同的颜色，并且沿着一条圆形路线首尾相连游动。离中心越近，它们变得越大。为了使每一行都与其周围形成鲜明对比，需要四种颜色。当所有这些成串的鱼像来自无穷远距离的火箭，以直角从边界射出并且再次落回到它射出的地方时，没有任何单个的成员到达边缘。因为边界是“绝对的虚无”。然而，如果没有其周围的空虚，这个圆形的世界也不会存在。这不仅仅是因为“内部”的先决条件是“外部”，而且还因为在这个“虚无”的外部，建立起这个框架的弧的中心点以几何的精确性被确定在那里。

这种描述可以说非常简洁地表达了庞加莱模型的实质。确实，埃舍尔的具有无穷数学魅力的“圆极限”以及它们的重复模式可以成为讲授双曲几何的有用工具。下面的右图是埃舍尔作品《圆极限Ⅳ》，其特征是魔鬼和天使交替出现。



第五节
非欧几何的影响

人类在第五公设的试证中经历了无数次的失败，最终却由于“一念转，天地新”，促成了非欧几何的诞生。在第一种非欧几何问世后，黎曼又提出了第二种非欧几何，从而发展了非欧几何。更重要的是，黎曼提出了广义的空间概念，建立了一种更为广泛的几何，于是两种非欧几何都可纳入内蕴微分几何的范围之内。其后，随着几种非欧几何模型的引入，非欧几何的相容性也得到解决。这使数学家相信，从逻辑角度而言，非欧几何和欧氏几何同样正确。于是，非欧几何终于在其创立者去世后获得了数学界的广泛认同。

非欧几何作为19世纪所有复杂的技术创造中最简单却最深刻的一个，它的创立不只表明第五公设是不可证明的（这解决了2000多年一直悬而未决的平行公设问题），更重要的是，它开启了众多新的问题，并在多方面产生了巨大的影响：它从根本上革新和增强了人们对几何学观念的认识，为几何学的发展开辟了崭新的途径，引出数学一些重要新分支；它使大家对几何学基础进行了深入研究，并最终从根本上使数学家们改变了对数学性质的理解；它还出人意料地应用于物理学，对空间观念进行了最深刻的革命；除对数学产生的巨大影响外，非欧几何蕴含的思想还对科学、哲学、宗教产生了革命性的冲击。正因此，人们有时把它与“哥白尼的日心说”“牛顿的引力定律”“达尔文的进化论”等相提并论。

几何学的统一

非欧几何的出现与被接受，打破了长期以来只有一种几何学（即欧几里得几何学）的局面，几何学从其传统的束缚中解放出来。当数学家们意识到存在其他不同的几何学时，他们开始自由地创造和研究他们自己发明的几何学。结果大批新的几何学（除罗氏几何、黎曼几何外，还有非阿基米德几何、非德沙格几何、非黎曼几何、有限几何等，还有与非欧几何并行发展的高维几何、射影几何、微分几何、拓扑学等）涌现出来，几何学这棵大树变得越来越茂盛。于是出现了这样的问题：什么是几何学？几何学是研究什么的？这些不同的几何学之间有什么关系？如何寻找不同几何学之间的内在联系，把这些不同的几何学统一起来？

在这方面，第一个重要的尝试由德国著名数学家 F. 克莱因提出。1872 年，年仅 23 岁的克莱因被聘为爱尔朗根大学的数学教授。按照惯例，他要做就职演讲。演讲中，他把几十年前代数学中刚出现的一种新思想——群——引入几何学。借助于群的观点，克莱因重新统一了几何学领域。

克莱因的基本观点是，几何学是研究几何图形对于某类变换群保持不变的性质的学问，每一种几何都由一个变换群所刻画，并且每种几何要做的就是考虑在这个特定变换群下的不变性质与不变量。例如欧氏几何由平移和旋转构成的变换群所刻画，它所研究的是长度、角度、面积等这些在平移和旋转下保持不变的性质。在这种观念下，19 世纪出现的几种重要的、表面上互不相干的几何学被联系到一起，而且变换群的任何一种分类也对应于几何学的一种分类。事实上，按照这种观点对几何学进行分类的话，欧氏几何、椭圆几何（包括单重与二重椭圆几何）、双曲几何等都包含于更一般的射影几何之中。

虽然并非所有的几何都能纳入克莱因的方案，但他的努力的确能给大部分的几何提供了一个系统的分类方法，因而其观点从根本上革新和增强了人们对几何学观念的认识，并对几何学的发展产生了持久的影响。克莱因的演讲及他提出的几何学研究规划，后来也以爱尔朗根纲领而闻名。

统一几何学的另一条途径为德国大数学家希尔伯特（1862—1943）所开通，这就是对现代数学影响深远的形式化公理方法。

正如我们已经介绍过的，公理化方法始于欧几里得的《几何原本》。在此书问世后的2000多年中，作为数学“圣经”，它一直被奉为数学严格性的典范。许多深刻伟大的数学家都竞相夸耀其“几何学的严格性”。到19世纪，随着数学中对更高“严格性”的倡导及非欧几何的诞生，数学家开始重新审视欧氏几何，发现《几何原本》的逻辑演绎体系确有不少漏洞和不够严密的地方。正如数学家罗素所评价的，“他的定义并非总是确定的，他的公理也不是都无法证明，他的论证需要许多公理，而他自己却没有意识到。严谨的证明应在没有图形辅助时依然保持其论证的力量，但欧几里得的许多早期证明却不能如此……他的著作作为逻辑名作的价值在很大程度上被夸大了”。鉴于此，人们对几何学基础进行了深入研究。

1899年，希尔伯特出版了名著《几何基础》，成功建立了希尔伯特公理体系。这一体系建立在基本概念和公理的基础之上。基本概念包括三个基本元素（点、直线、面）和三个基本关系（结合关系，分为点在线上与点在面上；顺序关系，指一点位于两点之间；合同关系，分为两线段合同与两角合同）。公理分五组：结合公理（8条）、顺序公理（4条）、合同公理（5条）、连续公理（2条）和平行公理（1条），共20条。希尔伯特公理体系的建立填补了欧氏几何的所有漏洞，使欧氏几何成为一个逻辑结构非常完善而严谨的几何体系，这标志着欧氏几何完善工作的终结。在这样组织起来的公理系统中，通过否定或者替换其中一条或几条公理，就可以得到相应的某种几何。例如用罗巴切夫斯基平行公理替代欧几里得平行公理，而保持其余所有公理不变，就可以

得到双曲型非欧几何的公理体系。如果在抛弃欧几里得平行公理的同时，添加任意两条直线都有一个公共点或至少有一个公共点的公理，并适当改变另外一些公理，就可分别得到单重与二重椭圆几何。这一工作，分别由数学家霍尔斯特德（1853—1922）与G. 赫森伯格（1874—1925）完成，前者在1904年建立了二重椭圆几何的公理体系，后者在1905年作出了单重椭圆几何的公理体系。至此，非欧几何有了严密的基础，而且有了统一的处理方法。

希尔伯特的工作，还标志着形式化公理体系的诞生。所谓形式化公理方法，是指在一个公理系统中，基本概念规定为不加定义的原始概念，它的含义、特征和范围不是先于公理而确定，而是由公理组隐含确定，即谁能满足公理组所要求的条件，谁就可作为该公理系统的基本概念、基本研究对象。也就是说，公理系统中的基本概念只具“形式”而不具“内容”，凡是符合公理组要求的对象都可以作为该“形式”的内容，公理组所阐述的是对基本概念的规定，而不是基本概念自明的特征。

因此，在希尔伯特公理体系中，几何学中“点”“线”“面”的直观意义都被抛弃了，研究的只是它们之间的关系，而关系由公理组来体现。即便有人把它们换成其他东西，只要符合公理组的要求，就都是可行的。正如希尔伯特一句名言所表明的：“我们必定可以用‘桌子、椅子、啤酒杯’来代替‘点、线、面’。”在这种处理下，原几何理论中所包含的特定意义（即对象）的直观背景被完全舍弃了。因为现在所从事的已不再是某种特定对象的研究，而只是由给定的公理（更准确地说是假设，因为现在已不具有特定的对象了）出发去进行演绎。因此，原有的那种“对象－公理－演绎”系统，变为“假设－演绎”系统。正因为这种系统不再具有明显的直观意义，因此称为“抽象的公理系统”。

希尔伯特的研究结果在1899年出版后，立即吸引了整个数学界的注意。形式的公理化研究方法很快被众多数学家所应用。现在，这种公理化方法几乎已渗透于数学的每一个领域，公理化结构亦成为现代数学的主要特征。

与这种更高度抽象的新研究方法相适应，希尔伯特还提出了具有更大普遍性的新研究问题，为数学开辟了新的研究领域。他所考虑的是一个公理体系建立后，如何判定其合理性。对此，他提出三条标准：相容性（或称无矛盾性、一致性），这是对公理体系最起码的要求；独立性，公理系中每一条公理不会是多余的，即其中任何一条公理都不能由这个公理系中其他公理推导出来；完备性（或称完全性），即系统中所有的定理都能由该系统的公理推出。

随后的问题是，如何证明公理体系满足这三方面要求。在《几何基础》一书中，希尔伯特对自己建立的几何学公理体系进行了这方面的研究。我们只简单提一下希尔伯特对最重要的相容性问题的处理方法。

我们上面已经介绍到，在希尔伯特之前，克莱因等数学家使用模型法（也称解释法）将非欧几何的无矛盾性归为欧氏几何的无矛盾性。这种方法是一种间接证明，其实质是将一个公理系统的无矛盾性证明归为另一个公理系统的无矛盾性。最初，人们相信把问题归结到欧氏几何的无矛盾性，问题已获解决。但希尔伯特认为，这还不够。他又向前进了一步，把解析几何看作实数系统中欧氏几何的一个解释模型，从而将欧氏几何的无矛盾性转化为实数系统的无矛盾性。当时实数系统已经可以建立在自然数的基础上，因此，实数系统的无矛盾性又归为自然数系统的无矛盾性。而对自然数系统无矛盾性的证明，又引出一系列新的数学思想，并发展出众多新的数学成果，极大地影响了 20 世纪数学的发展。追本溯源，这一切都与非欧几何的创立有关。

非欧几何不仅是数学史上的一座理论丰碑，而且还对人们的空间观念、哲学观念与数学观念产生了极其深远的影响。

观念革命

非欧几何建立后，它在逻辑上的无矛盾性由于各种模型的引入而得到证明。但还有一些紧密相关的问题，自非欧几何开始之时起，就受到数学家们的关注，即非欧几何可以描述物理空间吗？如果答案是肯定的，那么究竟哪种几何能描述我们所在的现实空间呢?

在非欧几何提出之前，欧氏几何的定理一直完美地与客观事实一致，因此几乎所有数学家都坚持物理空间的几何学，那唯一的几何学必是欧几里得式的。在近代科学中，欧氏几何成为牛顿绝对时空宇宙体系的数学模型。随着牛顿理论体系的成功，人们愈发坚定了对欧氏空间的信念。而绝大多数哲学家也都持有相同的看法，如 18 世纪的哲学大师康德断言：物质世界必然是欧几里得式的，欧氏几何是人类了解空间以及空间图形性质的唯一途径。可以说，直到 1800 年左右，所有数学家都认为欧氏几何是物质世界和此空间内的图形性质的正确理想化。这种认识由于康德等哲学家的阐发和影响更是牢牢地扎根在西方文化之中。

历史上，最早认识到“非欧几何能同欧氏几何一样用于表示物理空间”这一革命性思想的是数学之王高斯。在 1817 年的一封信中，他写道：“我越来越相信，欧氏几何的物理必然性是不能证明的，至少不能由人类理性也不适合由人类理性来证明。现在不行，也许在另一个世界中我们能获得对于空间实质的洞察。在到达那一步之前，我们不能将几何学放在纯粹先验的算术一类，而应该和力学放在一类。”最后一句是点睛之笔，在高斯看来，欧氏几何与力学定律一样都应建立在经验的基础之上。也就是说，欧氏几何并不是物理空间必然的几何学，其物理上的真理性并没有一个先验的基础来保证。这样，为判断哪种几何更适合我们所在的现实空间，只能通过实践来检验。

据一种流传很广的说法，高斯曾实际测量了以三个山峰为顶点的三角形内角和，试图检验哪种几何学更符合实际。从理论上来说，在欧氏几何中这个和恰是180°，而在罗氏新几何中这个和应小于180°。结果，高斯测得的三角形内角和与180°相差2″，考虑到实验误差的因素，所以这一不具判决性的实验什么也说明不了。

罗巴切夫斯基对这一问题也做了深入思考，并得出了同样的结论。他相信，不能不假思索地断言欧氏几何在我们所生活的世界里成立，“正如物理规律一样，它的真实性必须由诸如天文观测那样的实验来确证”。他猜测，非欧几何是“巨大尺度形式的几何”，即适合于大宇宙空间范围。他还进而推想，微观领域的几何也是非欧几何。他极富创见地向同代人宣告：“在观测不足的情况下，应当凭理智设想，想象几何适用于被观测到的世界之外以及分子引力范围之内。”为了证明自己的猜测，罗巴切夫斯基同样打算通过实验来确定哪一种几何学在现实中成立。在对由地球在它的轨道上两个对径点和天狼星所组成的三角形进行考察后，他发现该三角形的内角和与180°有一个小的偏差，但是这个偏差小于当时所允许的观察误差。因此，这一实验同样说明不了什么问题。

非欧几何另一创建人J.波尔约也探讨过究竟是欧氏几何代表“现实”还是非欧几何代表“现实”的问题，他的答案是：不能确定。事实上，他声称，“确定是否欧氏几何或某一非欧几何以及这个先验存在的不可能性留待证明”。

于是，当非欧几何创立时，三位非欧几何创立者都对传统观念提出了挑战，并都具有了革命性观点：非欧几何应用于物理空间是可能的，欧氏几何可能只是物理空间的近似写照。事实上，这正是高斯、罗巴切夫斯基、J.波尔约被称作非欧几何创立者的重要原因。

然而，非欧几何创立后的一段时间内，虽然数学家们慢慢接受了其逻辑上的合理性，但在很多数学家看来，只有欧几里得几何才能提供符合物理实在的空间模型，而非欧几何作为古怪的东西似乎与人们的经验认知很不一致，因此它与真实世界没有任

何关系。对此不必奇怪。对于从中学起就习惯了空间欧几里得解释的我们来说，不也是非常自然地认为我们的世界是欧氏的吗？一点也不令人吃惊。事实上，由日常经验支持的直觉使我们把欧几里得公理的正确性看成是理所当然的。

进一步促使数学界相信，物理空间的几何学可以是非欧几何的正是黎曼。黎曼在其著名的演讲中，不仅创造了可以与欧氏几何同样用来表示物理空间的黎曼几何，而且他宣称，物理空间的精确性质不是先验决定的，而只能由“经验”决定。在演讲的结束处，他还极富预见性地声称：“这条道路将把我们引到另一门科学领域，进入物理学的王国，进入现在的科学事实还不允许我们进入的地方。”黎曼的预言终于在20世纪为物理学的发展所证实。

1916年，伟大的物理学家爱因斯坦发表了广义相对论。在他的理论中，直线被定义为光线的路线。因为没有任何物质能够运行得比光速还快，所以在这种时空宇宙中光线代表了两点（“事件”）之间的最短路线。而且，这个理论预言光线在有强引力场的情况下应该弯曲。这意味着时空被弯曲了。但为了以数学方式描述这样一种弯曲空间，爱因斯坦曾徘徊了3年时间，最后他在大学同学数学家格罗斯曼的介绍下学习了在黎曼几何基础上发展起来的绝对微分学，亦即爱因斯坦后来所称的张量分析，并很快发现这正是建立广义相对论引力理论的合适数学工具。在1915年11月25日发表的一篇论文中，爱因斯坦终于导出广义协变的引力方程，从而宣告了广义相对论作为一种逻辑结构的大功告成。

简单回顾这一过程：首先是非欧几何提出了弯曲的空间，它为更广泛的黎曼几何的产生创立了前提，而黎曼几何后来又成为爱因斯坦广义相对论的数学工具。于是，我们看到了数学与物理有机结合的一则光辉典范。而作为历史上数学应用最精彩的例子之一，爱因斯坦的广义相对论反过来又恰恰揭示了非欧几何的现实意义。

随着广义相对论为一系列天文观测、实验所证实，人们开始确信，为其提供最恰当数学表述的黎曼几何看来比欧氏几何更适合刻画宇宙的结构。也就是说，描述宇宙

结构的几何学不是欧氏几何而是（广义）非欧几何的一种。当然，关于我们所在空间的几何选择问题是一个非常复杂的课题。实际上最重要的应该是，为了描述现实世界，应以怎样的精确度来选择适合于它的几何。如在地球范围内，空间用非欧几何来描述虽较为精确，但是这种非欧性极不重要，甚至完全可以把它放弃，因为这种非欧性用最现代化的测量仪器也不能发现。正因此，欧几里得几何在许多实际问题和工程技术等方面仍然保留着自身的全部意义。

剩下的问题是，在以黎曼几何为数学工具的广义相对论看来，我们的宇宙到底属于哪一种几何空间呢？

对此的研究建立在一个合理的假定“宇宙学原理”基础之上。这一原理是说，在任意时刻，在宇观尺度（10^8光年）上，宇宙中不同的空间点和同一点的不同方向上不存在任何差别，是均匀和各向同性的。注意，这一原理的合理性建立在“宇观尺度上”。也就是说，这里的“空间点”不是一般的“点”，其“半径”有10^8光年之大！正如我们所熟悉的，一个固体虽在原子尺度上物质分布不均匀，但在力学范围的宏观尺度上可认为是均匀和各向同性的，在较小的尺度上虽然宇宙中的物质分布是极不均匀的，然而当度量尺度放大到以10^8光年作为“空间点”时，完全可以认为宇宙中的物质是均匀和各向同性的。

由此假设出发，物理学家建立了标准宇宙模型，又称弗里得曼宇宙。于是，我们的问题变为，弗里得曼宇宙属于哪一种几何空间呢？可以证明，物质均匀和各向同性的空间必定是常曲率空间。我们已经介绍过，常曲率三维空间有三种：欧氏空间（曲率为0）、双曲空间（曲率为负）及椭圆空间（曲率为正）。于是弗里得曼宇宙只能是欧氏空间、双曲空间及椭圆空间这三种空间之一，而最终的答案取决于曲率的符号。进而由广义相对论可知，这一符号又取决于宇宙的真实平均密度ρ_c与理论上存在的一个临界密度ρ的关系。如果$\rho = \rho_c$，可以证明曲率为0，则弗里得曼宇宙是平坦的欧氏空间；如果$\rho < \rho_c$，可以证明曲率为负，则弗里得曼宇宙是双曲空间。这两种情况下的

宇宙都是开放的，其未来图景都是宇宙一直膨胀下去。如果 $\rho > \rho_c$，可以证明曲率为正，则弗里得曼宇宙是封闭的椭圆空间。这种情况下得到的是闭合宇宙，宇宙的未来图景是膨胀到一定程度停下来转为收缩。可惜的是，至今物理学家仍然无法确定出宇宙的平均密度，因此我们的宇宙到底属于哪一种几何空间现在还是一个谜。不过，我们所清楚的是，谜底不是先验决定的，而是由“经验”决定的。或如黎曼预言的，揭开这个谜需要物理学，而不是数学。

在彻底改变了人类的空间观念之外，非欧几何的建立还有力地冲击了人们的真理观。

在西方，2000 多年来数学曾是真理之堡垒，数学知识被普遍视为绝对不会错的知识，数学命题一直被看作是必然真理，而欧氏几何被认为是关于空间的绝对真理。这种执着的信念，在数学发展历史上，曾束缚了人们的思想，导致萨凯里等人未能踏进非欧几何的大门。然而，随着非欧几何的创立及应用，数学家们被迫面对这样一个问题：欧氏几何、罗氏几何与黎曼几何三种几何学不仅逻辑上都无矛盾，而且也都能用来描述现实的物理空间，那么它们之中哪一种几何学是自然界的真理呢？几种根本不同的几何学不能都真实，如果一种是真的，那另外两种与其矛盾的几何不应是假的吗？

对此的反思，使数学家认识到，数学空间与物理空间之间有着本质的区别，几何不是关于物理空间的真理，数学公理和定理并不一定是物理世界的真理。人们应放弃把真理与数学等同起来的传统观念。取而代之的新观念是，数学可以是来自经验启示的一种创造，其适用性是有限的，它在哪里适用只能由经验来决定。“数学并不是一堆天然的钻石，而不过是人工宝石”，人工的数学是人的思想的自由创造，而不是我们生活的世界派给我们的必然产物，它并不一定也不必提供真理性。于是，数学家只能满足于数学处于这样的地位：数学只是现实的一个相当有任意性的人为模型，对物理世界发生的事它只能或多或少地加以描述。人们不再宣称数学给出的是真理，“数学自命

为真理的态度必须抛弃了”，数学的真理性丧失了。自古希腊以来，每个时代都坚信存在绝对真理（欧氏几何就是一个典范）的信念破灭了，而人们对绝对真理的寻求也被证明是徒劳的。

除改变空间观，扫荡整个真理王国，导致真理观的巨大变化外，非欧几何的建立造成的另一重要转变体现在西方数学观，特别是数学研究思想上。

具体地说，非欧几何的确立使人们开始认识到，在以往的数学研究中人们往往不自觉地把直观和经验作为数学基础。这种以直观和经验作为数学基础的做法，带来两个方面的问题：首先，由于数学以直观和经验作为基础，数学与客观真理就被完全地等同起来——数学成了客观规律、真理体系的化身；其次，也正因如此，人们的思维和数学创造就被严格局限于直观和实践经验，从而就无法自由地展开思维创造的翅膀。

在这一意义上，非欧几何对欧氏几何的挑战也就是对以直觉和经验作为数学基础这一传统做法的挑战。为接受非欧几何的挑战，数学家们采取的做法是，取消数学对直觉与经验的依赖。正如我们上面提到的，在深刻影响后来数学发展的希尔伯特的形式化公理思想中，直觉的真理最终被消除。数学成为一个逻辑系统，它从一组真假值不确定的公设出发，只关心公理系统内部逻辑上的一致性，而不在乎系统与外在真实物理世界之间是否相符。数学变成一种毫无内容支撑的形式化建构，而有关外部世界的事实则被交给其他学科（主要是物理学）。

在这种数学观念下，我们可以说，欧氏几何与非欧几何都是真的。注意，这种真并不是指其在物理世界中的真假，而只是指几种不同的几何学在其各自的公理体系内，逻辑上不自相矛盾而已。换言之，其“真”并不是指物理的真实性，而是逻辑的真实性。如果进一步追问：不同几何学中哪种几何学的公理和公设是真的呢？对现代数学家来说，这样的问题是没有意义的。事实上，他们已不关心平行公设在物理世界是真还是假这一问题，而只关心它与其他公理的逻辑一致性。在他们看来，数学公理只是对数学对象性质的约定，它本身无所谓真假。

数学观念的这次革命，一方面把数学的地位由古希腊表现宇宙万物的构造下降为一种定义、公理和逻辑演绎的构造体系，这使数学丧失了真理性。然而，另一方面，这种开放式的数学观，使数学研究可以不受直观、经验的约束，从而扩展到没有明显的直观意义或现实背景的方向上。数学只需关注逻辑的相容性，而不必顾虑所创造的理论是否符合经验现实；数学不再束缚于直接从现实世界抽象而得的概念，而是有了去探索人类心智的创造自由。数学家能够而且应该探索任何可能的问题，探索和构建任何可能的公理体系。于是，这种全新的数学观，给整个数学界带来了自由的气息，为数学自身的发展开拓了无限广阔的前景，使数学获得了自由和新生。而数学也由此被引向了更广泛、更抽象、更理想化的理性创造领域。正是在这一意义上，非欧几何可被看成现代数学的实际起点。

结　　语

在本章中，我们介绍了人类在探求欧几里得第五公设问题时所走过的艰辛而漫长的道路。为了追根溯源，我们先是简单介绍了曾作为数学“圣经”的《几何原本》。在这本深刻影响了数学发展及西方文明的伟大著作中，欧几里得以五条公理与五条公设作为逻辑推演的前提，建立起一座巍峨的几何学大厦。然而，在这些作为起点的命题中，第五公设看上去却缺乏公设应有的简单性。于是，自欧几里得时代起，人们就以怀疑的目光审视这一公设，人们大都相信它事实上是一条可以被证明的命题。由此第五公设问题产生了，其后无数人踏上了第五公设的试证之路。

最初，人们希望能从欧几里得的其他公设和公理导出它，或者试图以更能接受的等价形式来取代它。然而，直到 18 世纪中叶，这样的尝试都失败了。每一种“证明”要么隐含着一个与第五公设等价的命题，要么存在着其他形式的推理错误。由于证明第五公设的尝试是如此之多，又都归为徒劳，结果法国数学家达朗贝尔（1717—1783）将第五公设问题称作“几何学基础中的丑闻”。

直接证明的长期失败启发人们改用间接证明方法。于是，我们看到人们在证明“第五公设”的漫长征程中，努力的方向开始发生了转变。许多数学家开始尝试用反证法，即从第五公设不成立的情况入手，追究能否由此得出与已知定理矛盾的结果。这种思想开辟了一条通往新几何——非欧几何——的道路，并由此涌现出几位新几何的先行者。历史表明，他们的工作孕育了非欧几何的诞生。

最早走上这条道路的是意大利数学家萨凯里，之后是瑞士数学家兰伯特。然而，当他们走到非欧几何的门槛前时，却在“只有一种几何学”的根深蒂固的信念支配下，或望而却步，或徘徊不前，未敢往前迈进。施韦卡特与陶里努斯沿着这条新道路又迈进了一步。在他们看来，欧几里得第五公设不能从欧几里得的其他 9 条公理、公设推导出来，也就是说，它独立于欧几里得的其他公理、公设。由此，他们相信存在其他逻辑上相容的几何，认识到了一种新几何存在的可能性。但他们都失去一个基本点，即欧几里得几何不是唯一的在经验能够证实的范围内来描述物理空间性质的几何。因为相信只有欧氏几何才是唯一能描述物理空间性质的几何学，他们未能突破具有 2000 年根基的欧氏几何传统的束缚，未能迈出决定性的一步。

“犹如紫罗兰在春季到处开放”，决定性的一步在 19 世纪由三位不同国籍的数学家各自独立迈出。

最早发展非欧几何理论并对非欧几何的意义有深刻理解的是数学之王高斯。然而由于担心受到世俗的攻击，这位行事谨慎的数学家决定将自己的发现秘而不宣。在高

斯之后，年轻的匈牙利数学家 J. 波尔约于 1823 年有了相同的发现，并在其“思想史上最不同凡响的二十几页”的论文中“从乌有中创造了一个新奇的世界”。然而，来自高斯的评语使 J. 波尔约十分失望，并基本放弃了这方面的研究。于是，发展并坚决捍卫非欧几何的重任落在了“几何学的哥伦布”俄国数学家罗巴切夫斯基身上。自 1826 年宣读自己第一篇关于非欧几何的论文起，直到去世为止，罗巴切夫斯基在孤境中以其非凡的勇气为非欧几何的生存和发展奋斗了长达 30 年，为非欧几何的创立做出了最重要的贡献。有心的读者应该能从三位非欧几何发明人对待自己发现的不同态度中得到一些有益启示。

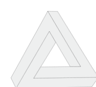
罗巴切夫斯基、高斯、J. 波尔约各自独立地提出了本质上相同的一种非欧几何，宣布了非欧几何的诞生。但非欧几何要获得普遍接受，还需要进一步确立自身的无矛盾性和现实意义。

1854 年，在自己的著名演讲中，德国数学家黎曼大大发展了非欧几何，建立起一种更广泛的几何，即现在所称的（广义）黎曼几何。罗氏几何（双曲几何）、欧氏几何、狭义的黎氏几何（椭圆几何）都成为这一（广义）黎曼几何的特例。而后，随着几种非欧几何模型的建立，非欧几何的无矛盾性也获得了解决。这些结果最终使非欧几何获得了数学界普遍的承认。这同时也对第五公设问题做出了判决：欧几里得第五公设不可能在其他 9 条公理、公设的基础上推导出来，它在欧氏几何中是一个独立的命题，因此人们可以换一个与之矛盾的公设并发展出全新的几何，如双曲几何与椭圆几何。

当回顾从试证第五公设到非欧几何诞生，再到非欧几何地位确定这一段漫长的历史时，我们可以清楚地意识到，事情的转折并非来自“技术性内容”。事实上，在非欧几何创立之前，已经有许多数学家发展了这些“技术性内容”。因此，在非欧几何的历史上，最具根本性的是思想观念的转变。也正因为观念的更新，非欧几何诞生了。而正因此，由观念更新引出的这门数学分支在其地位确定后，反过来在数学中产生了深

刻而巨大的影响。通过上面的介绍，我们已看到，非欧几何不仅成为数学史上的一座理论丰碑，而且还对人们的空间观念、哲学观念与数学观念产生了极其深远的影响。从探求第五公设问题的解答到孵出非欧几何这个数学金蛋，到非欧几何对数学的深刻影响，细细品味这段数学史，真能获益良多。





第四章

四色问题

第一节
初识四色猜想

在绘制地图时，为了区别不同的区域，通常要对任意两个有共同边界的相邻区域染上不同的颜色，而为了节省颜色，对不相邻的区域则不必涂不同的颜色。你是否留意过，按这种要求绘出的地图实际上至多用了 4 种不同的颜色？当有人灵机一动，想到“是否任何地图都可以用 4 种颜色来染色”时，世界最迷人的数学难题之一——四色问题问世了。

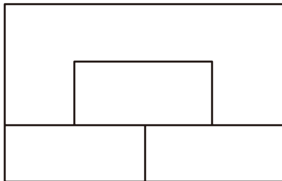
四色问题的来源

1852 年，一位从伦敦大学毕业名叫弗兰西斯・格思里（1831—1899）的学生，在研究英国地图时注意到一个有趣的事实：英国的郡都可以用 4 种颜色来染色，且使得相邻的郡染不同的颜色。在进一步研究中，他发现有些地图用 3 种颜色无法完成染色，但似乎所有地图用 4 种颜色就足够了。具有数学头脑的格思里想到一个数学问题：是否所有地图都可以仅用 4 种颜色染色呢？他猜测答案是肯定的。这就是著名的四色猜想。格思里当时

还无法预料到这一问题究竟有多难，更无法预料到他会以这一问题的提出者而留名于数学史册。

不久后，格思里把自己的这个发现写信告诉了仍在伦敦大学读书的弟弟弗里德利克·格思里，并且画了一个图，这个图最少要用 4 种颜色，才能把相邻的部分区分开。恰好弗里德利克是著名数学家德·摩根的学生。于是在 1852 年 10 月 23 日，他把哥哥的问题与想法转交给德·摩根。

对此非常感兴趣的德·摩根在同一天写信把问题告诉了著名数学家哈密顿。在去信中德·摩根写道："我的一名学生今天请我给他解释一个事实，但我并不知道那究竟是不是一个事实——并且到现在还不知道。他说如果任意划分一个图形并给各部分着上颜色，使任何具有公共边界的部分颜色不同，那么需要 4 种颜色就够了，无须更多的颜色。下图是需要 4 种颜色的例子。

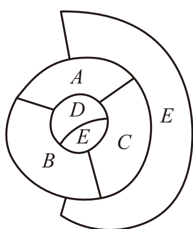


现在的问题是是否会出现需要 5 种或更多种颜色的情形……此事如果当真，难道从未有人注意过吗？我的学生说这是在给一幅英国地图着色时提出的猜测。我越想越觉得这是显然的事情。如果您能举出一个简单的反例来，说明我像一头蠢驴，那我只好重蹈斯芬克斯的覆辙了……"

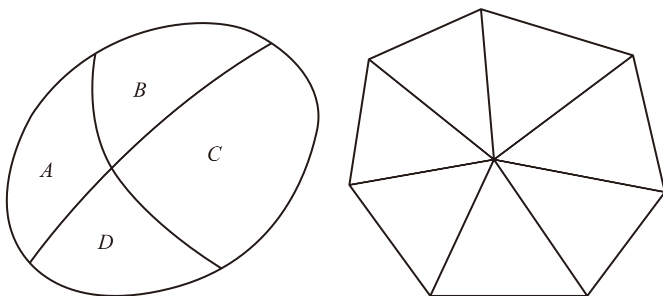
为了避免误解，我们先对四色问题——给任何一张平面地图着色，并对任何具有共同边界的国家（区域）着不同的颜色，只用 4 种颜色就够了——做几点说明。

首先，地图中任何一个国家都必须是连通的，连通的意思是说从国家内的任一点出发，都可以不出边界，走到该国家内任一点。简言之，即不能出现一个国家在地图

上由两个或更多的地区组成的情况，比如下图所示的地图就不符合要求，因为被分成两块的 E 是不连通的。



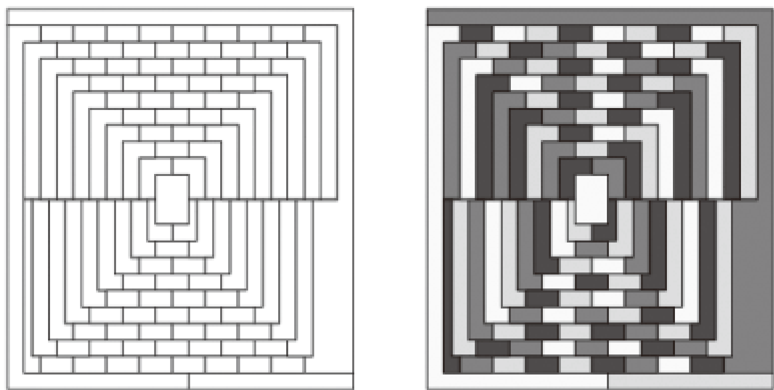
其次，问题中的相邻区域是指有一整段共同边界，而不是只有一点（或几个点）相连的区域。比如在下面的左图中，A 与 B、D 是相邻的，但 A 与 C 有一点相连就不能看作相邻区域。因此，A、C 可以用相同的颜色。否则，如下面的右图所示，各个国家像馅饼那样“相邻”排列，则有多少国家地图就会需要多少种颜色，四色猜想显然是不成立的。



此外，作为一个数学问题，我们要讨论研究的不是哪一张具体的地图，而是要考虑任意一张地图。比如，1975 年 4 月 1 日《科学美国人》杂志上，著名的数学专栏作家马丁·加德纳在一篇文章中煞有介事地宣称，下面这张地图来自一位图论专家发给专业杂志的论文，它不能用 4 种颜色着色！如果有人按照染色要求，给出了它的一种四着色，是否意味着已经证明了四色问题呢？答案自然是否定的，这只能得出加德纳的宣称是错误的。事实上，这篇发表在愚人节的文章本身就是加德纳开的一个玩笑。

顺便指出，用 4 种颜色对一个复杂的地图染色并不是一件容易的事情。据说，加

德纳的这个由 110 个区域组成的地图曾经被美国一些大学选作计算机或数学专业学生的竞赛题或习题以征求答案。但长时期中见不到该图的具体四着色方案。这里给出了一种解答，如下面的右图所示。



收到德·摩根来信的哈密顿并没有表示出对此问题的浓厚兴趣。在 10 月 26 日的回信中，哈密顿写道："我可能不会很快就考虑你的颜色'4 元组'问题。"

对此深感兴趣的德·摩根在随后的几年里继续研究这一问题。1860 年 4 月 14 日，在一本杂志上评价一本《发现的哲学》的书时，德·摩根再次提到四色问题，并谈了自己对它的看法，但由于一直未获大的进展，他后来对这一问题的热情也慢慢减少了。正当这一问题在人们的视野中逐渐淡出时，另一位英国数学家凯莱（1821—1895）对此的关注复燃了人们对它的兴趣。

1878 年 7 月 13 日，在伦敦数学学会的一次会议上，凯莱询问四色问题是否已经得到了证明。在得到否定的回答后，凯莱会后发表了一篇《论地图着色》的文章。文章中他更严格地表述了这一问题，并指出这个问题的一些困难所在，最后他承认自己一直没有得到这个问题的严格证明。

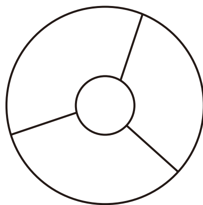


凯莱

四色问题在经过凯莱正式提出后，引起了人们普遍的关注，并在当时掀起了一场四色问题热，许多人加入到证明四色猜想的队伍中。在介绍其他人的工作之前，我们先来看一下德·摩根取得的小进展。

德·摩根的工作

首先，德·摩根与格思里一样，一开始就认识到在为某些地图染色时 3 种颜色是不够的，这表明要用不同的颜色区分不同的区域至少要 4 种颜色。下图即为一个不能用 3 种颜色染色的简单例子。



此外，德·摩根还得到了一个涉及邻域的结论。有关邻域的问题，最早出现在德国著名数学家茂比乌斯和他的一位朋友的讨论中。1840 年，在一次演讲中，茂比乌斯第一次明确提出这个问题，并讲了一个有趣的童话。

从前，在远东生活着一个国王，他有 5 个儿子，他们将在国王驾崩后继承王位。国王在遗嘱中规定，他的王国将分成 5 部分，每一个都必须与其他 4 个毗邻。因为老国王担心，假如不这样分割，一个兄弟去访问另一个就不得不经过他人的国土，兄弟就可能变得疏远。另外，国王还规定，每对兄弟必须修筑连接他们土地的马路，而这些道路必须分开，不能穿过第三者的领地。国王死后，五弟兄费了很长时间来遵照父亲的遗愿分土地，但他们的努力都失败了。

弟兄们完不成的这项任务，暗示了平面上 5 个区域中的每一个区域都与其他 4 个区域相邻是不可能的。进而言之，平面上相互毗邻区域的最大数目，等于平面上能用互不相交的简单曲线连接的点的最大数目，而这个最大数目都等于 4。这正是德·摩根后来所证明的结论。

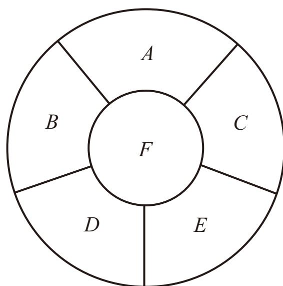
这一不难发现并证明的事实，为许多试图证明四色猜想的人提供了一个陷阱。许多人误以为这一结论已蕴含了四色定理，并从此事实出发推得四色猜想成立。我们分析一下这种推理错在什么地方。

假设平面上存在 5 个区域，其中每一个区域都与其他 4 个区域相邻，那么显然为这 5 个区域着色需要 5 种颜色，因而四色定理不成立。正如德·摩根证明的，事实上这一假设不成立。然而由它的不成立，能否直接推出四色定理成立呢？当然不能。因

为，“没有地图包含 5 个两两相邻的区域”这一结论并不意味着，不存在其他需要 5 种颜色的地图。

为了说明这种蕴含关系不成立，我们还可以采用以毒攻毒的方法。

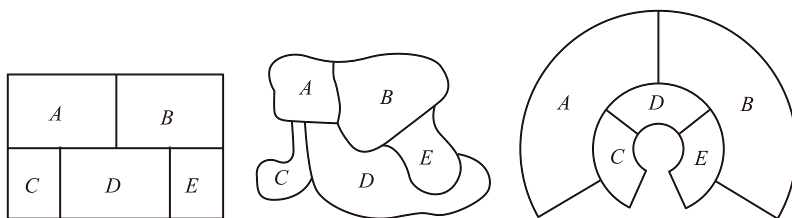
如果认为“任何地图都不允许其中 5 个国家都与其他 4 个相邻，则用 4 种颜色足以给地图着色”成立，那么类似的应该有“任何地图都不允许其中 4 个国家都与其他 3 个相邻，则用 3 种颜色足以给地图着色”成立。但对后者，我们可以举出下面的反例图。



如图所示，在这 6 个区域中，根本没有 4 个区域是每个都与其他 3 个相邻的，但为了给这幅地图染色，3 种颜色并不够，事实上它需要 4 种颜色。

由此，我们看到，地图着色所需的色数并不一定等于彼此相邻国家的最大数目。也就是说，虽然证明了平面上不会有 5 个彼此都相邻的区域，然而这并不能成为四色定理的证明。在后面的介绍中，我们会看到许多关于四色猜想的证明，因为落入不同的陷阱而碰了壁。这里提到的正是四色猜想者容易遇到的第一块暗礁，也是刚入门者必须小心的地方。事实上，一些号称证明了四色猜想的人就是碰在了这块小暗礁上却不自知。

对我们要探讨的四色问题来说，还有一个比较明显但重要的事实：问题中各个区域的实际形状与大小并不重要，重要的仅仅是它们的相对位置。比如，下面三幅不同的地图，在考虑着色时是完全等价的。



这种等价在数学上称为拓扑等价，由此进入的是一门称为拓扑学的数学分支。下面我们暂且放下四色问题，去简单了解一下拓扑学，一门起源于游戏的趣味数学学科。

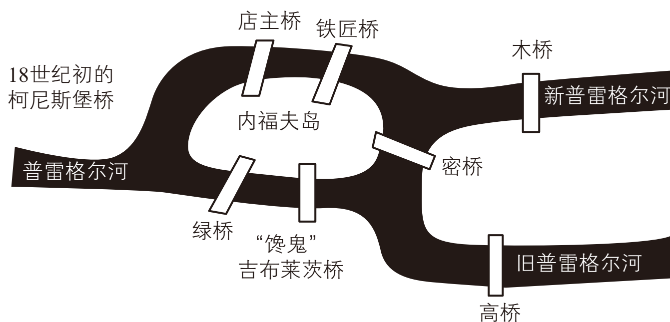
第二节
拓扑学与图论：起源于游戏的数学

数学游戏，作为一种运用数学知识的大众化智力娱乐活动，因其趣味性，往往容易激发人们的数学热情与兴趣，因而对普及数学而言是非常有益的。另一方面，从历史角度看，数学游戏还曾对数学的发展起到过重要而积极的作用。特别是，它曾直接促进新的数学分支的创立。我们这里要介绍的拓扑学与图论就是典型的例子。

柯尼斯堡七桥问题

柯尼斯堡城建于1255年，原是德国东普鲁士的一部分，第二次世界大战后，划归苏联（现属俄罗斯），并被更名为加里宁格勒。这座名城在数学上经常被提起，则源于与它有关的一个著名数学问题。

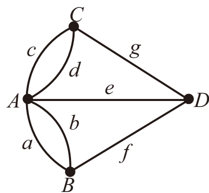
这一问题远溯至18世纪初。当时有一条名为普雷格尔的河从柯尼斯堡城中穿过，河中有两个岛，河上有7座桥连接着这两个岛及河的两岸，如下图所示。



柯尼斯堡城的居民喜欢沿着城市的河岸和岛屿散步。不知从何时起出现了一个有趣的娱乐问题：能否找到一条路线，可以经过所有 7 座桥，但不重复经过任意一座桥（而且回到起点）？

这一问题后来传到了伟大数学家欧拉耳朵里，并引发了他的研究兴趣。1736 年，欧拉向俄国科学院提交了一篇论文，彻底解决了这一问题。

欧拉先是把问题做了简化处理。他用点代表陆地，用线段或弧代表桥，具体点说就是用 A、D 表示两个小岛，点 B、C 表示河的左右两岸，再用连接两点的线表示桥，这样就把原图变成了一个由 4 个点和 7 条线组成的新图形。



在经过这样的抽象后，原来的问题相当于，能否用一笔画出上面的图。于是，“柯尼斯堡七桥问题”转化成“一笔画”问题。如果上图能一笔画成，就说明“一次连续走过七桥而无重复”的走法是存在的；反之，如果能证明不可能一笔画出上图，就说明这种走法是不存在的。

随后，欧拉先研究了一个更一般的问题：什么样的图形能一笔画出。

要想一笔画出一个图，总要选一个起点，选一个终点，还要经过一些中间点。也就是说，一笔画的过程是点、线相间排成一连串：起点－线－顶点－线……线－顶点－线－终点。起点可以由几条线汇合，但是画图时，总是先从它出去，然后进来出去几次（进出一次，得到两条线：进来是一条线，出去也是一条线），而最后一次是出去的，所以集中在起点的线只能是一条、三条、五条……即是奇数条。终点是先画进去，然后出去进来几次，而最后一次是进来的，所以集中在终点的线也只能是奇数条（这样的点我们称为奇点）。作为中间各点，有一条“进入线”就有相应的一条“离开线”，应是进去出来的次数相等，即每一点上都只能有偶数条线（这样的点我们称为偶点）。因此简单地说，中间点一定是偶点。于是，如果起点与终点是不同的两个点，则只有起点和终点是奇点，共两个奇点；如果起点与终点重合，即最后又画回到起点，那么起点、终点也成了偶点，这时所有的点就都是偶点了，即奇点数为 0。

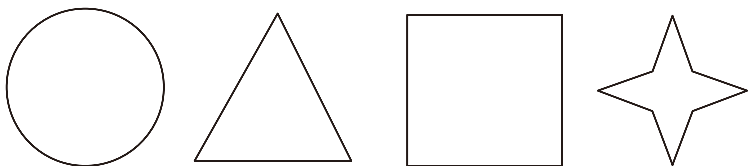
由此，我们可以得出一个结论：一个图能否一笔画出，是由它的奇点个数来决定的。能一笔画的图形，其奇点的个数只能是 0 或 2。

最后将这个结论应用到具体的七桥问题上。如果要不重复地走完柯尼斯堡的七座桥，那么转化后的图中奇点个数只有为 0 或 2 时才行，如果加上回到起点的要求，那么奇点个数只能为 0。但可以发现，上图中的奇点个数有 4 个，因此可以下结论说，这样的散步路线是不存在的。

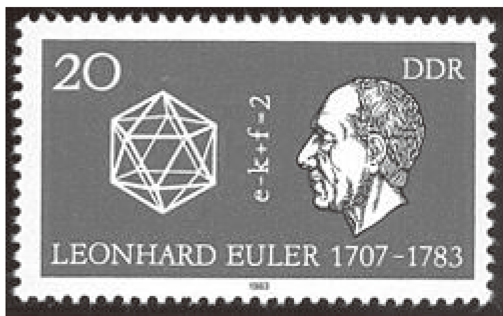
就这样，欧拉通过抽象分析思考把七桥问题归为“一笔画问题”，并以令人难以置信的轻巧解决了一笔画的可能性问题，从而圆满解决了柯尼斯堡七桥问题。对数学而言，欧拉这一颇受赞誉的研究更重要的贡献在于它拓宽了数学的研究领域。

在论文的开头欧拉曾写道：“讨论长短大小的几何学分支，一直被人们热心研究着。但是，还有一个至今几乎完全没有探索过的分支，莱布尼茨最先提过它并称之为‘位置几何学’。这个几何学分支只讨论与位置有关的关系，研究位置的性质，不考虑长短大小，也不涉及量的计算。但是至今未有过令人满意的定义，来解释位置几何学

的课题和方法……”欧拉在文章中提到的这门新几何学分支现在被称为拓扑学。正如欧拉所提到的，在这门新几何学中，我们平常所熟悉的距离、大小是无关紧要的，直线、圆、角度等也失去了意义。它所关心的是图形在弯曲、拉伸、压缩或扭转等连续变换下保持不变的拓扑性质。正因为研究这类性质，所以拓扑学获得了一个更为通俗的叫法“橡皮几何学”。在拓扑学中，一个图形经过弯曲、拉伸等拓扑变换得到的图形与原图形是拓扑等价的。如下图所示，几个图形从拓扑学角度而言就是等价的。所以，有数学家幽默地评论说：“拓扑学家是一群无法区分甜甜圈与咖啡杯的数学家。”顺便指出，可以证明平面上的地图与球面上的地图是拓扑等价的。因此，在球面地图上也有与平面地图等价的四色猜想。



欧拉对七桥问题的圆满解决，成为拓扑学的先声。在拓扑学上，欧拉做出的另一重要贡献是发现并证明了简单多面体的欧拉公式：$V-E+F=2$，其中 V 是顶点的个数，E 是棱的个数，F 是面的个数。为纪念欧拉的这一重要发现，德意志民主共和国还曾发行了如下邮票。



此外，欧拉对“七桥问题”的研究还奠定了另一门新兴的有趣数学分支——图论的基础。他在解决柯尼斯堡七桥问题时所得出的关于一笔画的结论也成为图论历史上

的第一个成果。对此，后来对图论做出大量重要贡献的数学家塔特写下一篇反映图论的幽默诗作：

> 柯尼斯堡的一些市民，
> 漫步在河畔。
> 在普雷格尔河旁，
> 有七座桥相伴。
> “欧拉，我们一起散步吧！”
> 那些市民在召唤。
> “我们在这七座桥上漫步，
> 仅经过每座桥一次。”
> “你们做不到，”欧拉大声说道，
> “结果就是这样，
> 岛屿作为顶点，四个点有奇数度。”
> 从柯尼斯堡到 König 的书，
> 图论的传说正是如此，
> 而且越来越精彩，
> 绽放在密歇根和耶鲁。

有趣的是，图论起源后的下一次进展也来自一个数学游戏，而游戏的发明者则是我们前面已提到的哈密顿。

神童哈密顿

哈密顿（1805—1865），爱尔兰数学家，数学史上诸多早熟的天才之一。他于1805年出生在都柏林。当他大约一周岁时，父母把他交给一位叔父抚养并进行教育。这位叔父是个语言学家，在他的培养下，天赋超群的哈密顿3岁时就能顺利地读英文，5岁就能读能译拉丁文、希腊文和希伯来文，8岁时又掌握了意大利文和法文，不到10岁就学阿拉伯文和梵文。14岁时，他用波斯文写信给当时正在访问都柏林的波斯大使。15岁那年，一件偶然的事件改变了他的兴趣与生活道路。当时，一个美国青年科尔布恩在都柏林表演了他的快如闪电的计算能力。“后来很长一段时间，”哈密顿后来写道，“我喜欢心算很长的算术运算，开平方和开立方，以及每一件与数的性质有关的东西。”最后，他决心终身从事数学。他宣称：“再也没有比研究科学更能升华人的心灵，更能把一个人提升到众人之上了。谁不愿意享有阿基米德的名声呢……”

1823年，这位神童进入都柏林的三一学院。21岁时，他提交了一篇重要的有关光学的论文。1826年，尚未大学毕业的哈密顿任三一学院天文学教授职位。1835年，哈密顿获爵士称号，两年后又当选为爱尔兰皇家科学院主席。

事业上的成功却没有给他带来个人生活的幸福。他于1833年结婚，妻子“极为羞怯，又极为纤弱”。在这一不幸的婚姻中，哈密顿夫人难以胜任家务且常常生病，结果导致悲惨的哈密顿食无定时。当哈密顿去世后，人们看到他的工作房间因长期无人打扫而变得如猪圈一般：房间里堆满了像小山样的纸张，而纸堆中竟然夹杂着大量的盘碟与剩下的饭菜。

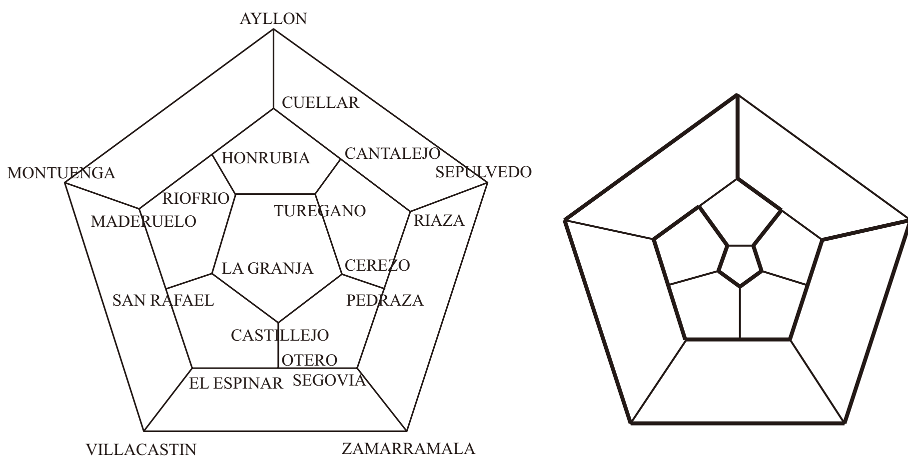


哈密顿

哈密顿在数学方面主要研究代数，其最重要的贡献是在 1843 年发现了四元数。把他与图论联系在一起的是他提出的一个趣味问题：从正十二面体（共有 20 个顶点）的一个顶点出发，沿着边行进，把 20 个顶点无遗漏地全部通过，每一个顶点只许过一次而回到原处，问这种走法是否存在。将这一问题略加转化，哈密顿于 1857 年发明了著名的周游世界游戏，并以 25 英镑把这个游戏卖给一家公司。后来，这一游戏还进入了市场。

游戏的道具由一个木制的正十二面体构成，正十二面体的每个棱角上标有一个当时非常有名的城市。游戏的目的是环绕世界旅行，其要求是寻找一个环游路线，使得沿正十二面体的棱，从一个“城市”出发，经过每个城市恰好一次，最后回到原出发点，并且经过的棱不许重复。

为了容易记住哪些城市已被游览过，在每个棱角上放一个钉子，再用一根线绕在那些旅游过的城市（钉子）上，由此获得旅程的一个直观表示。后来，人们发现这个道具不好用，因而诞生了与正十二面体等价的平面木板状版本。可惜，这个游戏的商业运作是失败的，玩具商没有从这一游戏中赚到钱。后面的右图给出了这个游戏的一种正确玩法。



让我们对照一下欧拉与哈密顿研究的问题。欧拉要求在一个图中寻找满足“从一个顶点出发，沿边行进，无遗漏走遍所有的边，每一边只许经过一次而回到出发点”的路线，这样的路线后来在图论中被称为欧拉回路。而具有欧拉回路的图称为欧拉图。与之表面相似，实际完全不同的哈密顿问题则要求在一个图中寻找满足“从一个顶点出发，沿边行进，无遗漏地通过全部顶点，每一顶点只许过一次而回到出发点”的路线，这样的路线后来在图论中被称为哈密顿圈，而具有哈密顿圈的图被称为哈密顿图或说图是哈密顿的。

欧拉问题与哈密顿问题已经成为现在图论的重要组成部分。令人感到惊奇的是，对于欧拉问题，我们已经有了简单、漂亮的结果：一个连通图存在欧拉回路当且仅当它的所有顶点都是偶顶点。然而对哈密顿问题即“一个给定的连通图是否存在哈密顿回路”，至今仍然是图论中一个尚未解决的著名难题。数学家们经过努力得到了一些存在哈密顿回路的必要条件与充分条件，但至今还没有得到充要条件。

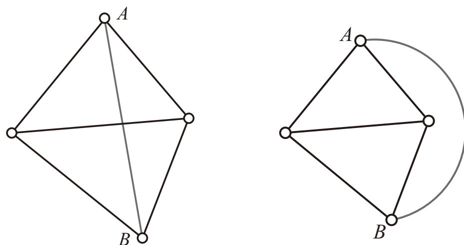
欧拉七桥问题与哈密顿周游世界游戏是两例标志性建筑，它们播下了图论诞生与发展的种子。这门一开始以游戏形式出现，而且此后也一直没有完全失去这一特点的数学分支，在20世纪得到迅速发展，已经成为十分有用的重要数学分支。

图论以图为研究对象。所谓图，就是一些点与某些点之间的连线（称为边）的集合。这里，我们仅介绍其中几个简单的名词。

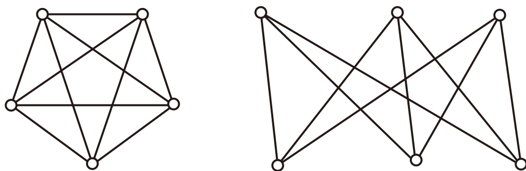
如果 $e=uv$ 是图的边，则称 u 和 v 是邻接的顶点，顶点 u 和 v 互称为邻点。一个点的相邻点的个数称为这个点的次数（或度）。

如果在一个图中，每一个顶点都可以通过边到达任何其他顶点，我们就称这个图是连通图。

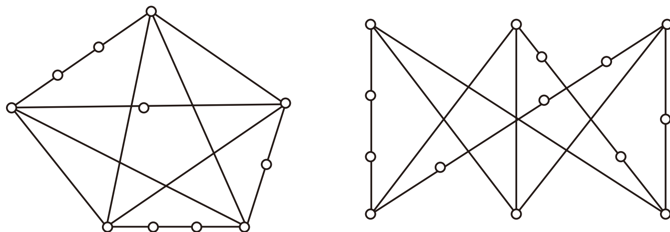
如果一个图可以画在平面上使得其中各条连线彼此不相交，那么这个图叫作平面图。当然，有些图（如下面的左图）初看起来不是平面图，但经过适当的移动（将 AB 边拉出来，从外面走），就可以变成没有相交连线的平面图，这种图称为可平面化图，或者称为可以嵌入到平面的图。移动前后两个图的点数相同，边数相同，点与边的关系也相同，它们称为两个同构图。



有的图无论怎么做都不能变成平面图，这种图称为不可平面化的图。最典型的不可平面化的图一个称为 K_5，一个称为 $K_{3,3}$，如下图所示。



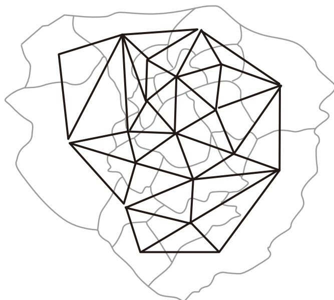
1930 年，波兰数学家库拉托夫斯基得出一个著名结论：图 G 可平面化当且仅当 G 不包含 K_5 或 $K_{3,3}$ 的剖分图。所谓剖分图，就是在原图的边上任意添加一些顶点所得到的图。



四色问题曾在近代图论发展中起过重要作用，那么这个地图染色问题是如何与图论联系起来的呢？为建立起图论与地图着色问题之间的联系，数学家引入了对偶图的概念。

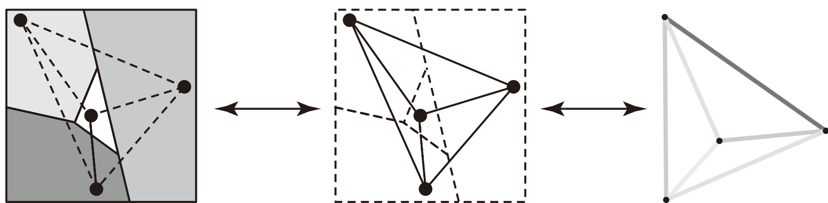
对偶图

四色猜想是一个区域的染色问题，它可以看作是图的面染色。此问题可以换一种角度考虑：在地图的每一个区域内取一点，为了形象，不妨设想这点是区域的首都，然后将有公共边界国家的首都用线连接，由此得到一个由点与线构成的新图。显然，每张地图都可以通过这种方式得到一个与它关联的新图，我们称其为原地图的对偶图。图中的点称为顶点，图中的线称为边。图中由边相连的两点称为相邻点。显然，一地图中两个顶点如果是相邻点，那么当且仅当它们所对应的区域有公共边界。



地图的对偶图具有一个特点：其边除了可能在顶点上相交外，都无其他交点。用我们上面刚引入的术语来说，就是每张地图的对偶图都是平面图。反之，每个连通的平面图都是某个地图的对偶图。于是，经过这种转化后，地图中的面着色问题等价于平面图中的顶点着色问题，而“有公共边界的国家着不同的颜色”变为“每条边的两个端点着不同的颜色”。而四色猜想在其对偶图中等价于，对一个平面图上的顶点染色，使相邻两顶点颜色相异，至多用 4 种颜色就够了。如果把平面图顶点染色所需要的最少颜色数称为平面图的色数，则四色猜想也等价于，每个平面图的色数至多是 4。

在引入对偶图后，我们还可以做另一种转换。考虑对所得的平面图进行边染色。即让一个图的每边染一种颜色，并使有公共顶点的边（称为相邻边）有不同的颜色，这样的染色称为边染色。同样，如果把平面图边染色所需要的最少颜色数称为边色数，则四色猜想等价于，每个平面图的边色数至多是 4。



这样，四色问题被带进了图论领域，并转化为图论的着色问题。在下一节中，我们将看到一种通过研究图的着色来解决四色问题的重要尝试。

第三节
捷报频传

1878 年，英国当时最著名的数学家凯莱正式向伦敦数学学会提出了四色问题。这一表述简单的问题立即吸引了人们，众多数学家与数学爱好者纷纷加入到四色猜想的大会战中。仅仅一年后，好消息传来，一个叫肯普的人宣布自己证明了四色猜想。又过了一年，捷报再传，一个叫泰特的人宣布找到了证明猜想的另一种途径。

震动数学界的 8 页论文

在凯莱重提四色猜想的那次会议上，有一个名叫肯普的与会者。肯普（1849—1922）于 1872 年毕业于剑桥大学，曾跟随凯莱学习数学。后来成为一名律师，但他一直对数学感兴趣。在听到四色问题后，肯普马上投入到研究中。1879 年 7 月 17 日，肯普在《自然》杂志上宣布他已经解决了四色问题。在凯莱的建议下，肯普把自己的论文发表于《美国数学杂志》第二卷上。这篇仅 8 页的论文震动了数学界。许多数学家对他的发现给予了

高度的评价，肯普也因自己的这一成果获得了很大荣誉。1881 年，他当选为英国皇家学会会员，后来还因对数学的贡献被封为爵士。



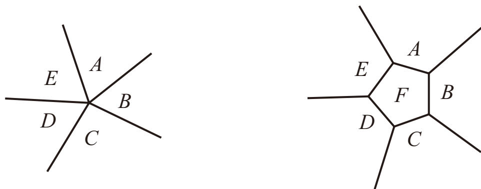
肯普

肯普的方法极为巧妙，并且引入了大量基本的思想。下面我们就来欣赏一下他的证明思路。

正规地图

肯普首先引入正规地图的概念。正规地图是指满足如下条件的地图：其中没有一个区域包围其他区域，也没有 3 个以上的区域相遇于一点。非正规地图总可以转化为正规地图。比如，出现 3 个以上的区域相遇于一点时，可做如下处理：在交点处加入一个新的区域，使某些不接触的区域变成接触。以下面 5 个区域相交于一点的左图为

例，只需要在交点处添加一个新的区域 F，于是在得到的新地图中有 5 个交点，每个都仅是 3 个区域的交点。而且易知，若新的正规地图可以用 M 种颜色着色，则把区域 F 缩小成一点后，就可得到原地图的不多于 M 种颜色的着色方法。



由于非正规地图染色所用的颜色数不超过处理后得到的正规地图所用的颜色数，因此，如果所有正规地图能用四色染，那么非正规地图用四色染也足够了。于是，对于四色问题，我们可以只考虑它在正规地图中是否成立。

任何一张地图的四色问题都可变成一张所谓正规地图的四色问题，这种处理在数学中是很有用的，它不仅使问题得以简化而且使问题变得统一。如果把正规地图转换为它的对偶图，这种统一性就更加明显：在正规地图的对偶图中，各个面都是三角形。简化、统一带来的好处是，它使问题在数学上容易处理。比如可以证明，任何正规地图中都至少存在一个国家，其邻国数少于 6 个（即至少包含一个边数少于 6 的国家）。下面简单解释一下。

设地图中共有 F 个区域，V 个顶点，E 条边，并设具有 2 条边界的区域有 f_2 个，3 条边界的区域有 f_3 个…… n 条边界的区域有 f_n 个。于是，$F = f_2 + f_3 + f_4 + f_5 + \cdots + f_n$。由于每个 f_2 这类区域有两条边界，所以这类国家共有 $2f_2$ 条边界。同理，f_3 类国家共有 $3f_3$ 条边界……。又由于每条边属于两个区域，所以边界总数 E 满足 $2E = 2f_2 + 3f_3 + 4f_4 + \cdots$。又因为在正规地图中，每个顶点属于三个区域，而每条边属于两个区域，因此有 $3V = 2E$。

根据欧拉公式 $V + F - E = 2$，消去 V，可得 $6F - 2E = 12$。

于是，$6\times(f_2+f_3+f_4+f_5+\cdots+f_n)-(2f_2+3f_3+4f_4+\cdots)=12$

即，$(6-2)f_2+(6-3)f_3+(6-4)f_4+(6-5)f_5+(6-6)f_6+(6-7)f_7+\cdots=12$

亦即，$(6-2)f_2+(6-3)f_3+(6-4)f_4+(6-5)f_5=12+f_7+\cdots$

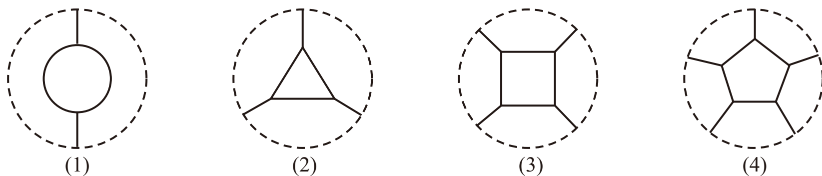
因为这一等式右端大于零，所以等式左端的f_2、f_3、f_4、f_5不能同时为零。这表明，一定存在边界数少于 6 的区域，即任何正规地图中都至少存在一个国家，其邻国数少于 6 个。

在提出正规地图的概念后，肯普下一步要做的是证明任何正规地图都不需要用 5 种颜色染色。为此，肯普使用了一连串反证法。其思路是，假设存在正规的五色地图，那么在这些正规的五色地图中，自然存在一个国家数最少的“最小正规五色地图”（换言之，这种最小正规五色地图需要 5 种颜色着色，但任何国家数少于它的地图只需 4 种颜色着色），然后只需推出这种“最小正规五色地图”不存在即可。为此，肯普又假设“最小正规五色地图”存在，他打算推出存在比它国数更少的正规五色地图，从而导出矛盾。这样最终得到正规地图的四色定理。为了推出存在比“最小正规五色地图”国数更少的正规五色地图，肯普又引进了一些重要概念与思想。

不可避免的可约构形集

肯普首先证明了我们上面提到的结论：在任何正规地图中，都不可能存在所有国家都有 6 个或 6 个以上邻国。换句话说，任何正规地图中都必存在一国具有 2、3、4

或 5 个邻国。这就意味着，所有正规地图至少含有下列 4 种构形（构形是指地图中一组连接着的区域，以及每个区域外边有多少个区域与它相邻的信息）的一种，其中中间的国家分别与 2、3、4、5 个国家相邻。



用一个专业术语来说，那就是由 2、3、4 或 5 个邻国组成的这些“构形”的集合是不可避免的构形集合。

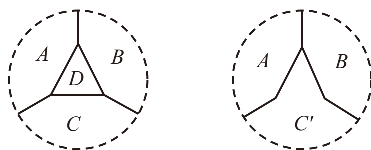
肯普进而要做的是，在存在“最小正规五色地图”的假设下，从不可避免的构形集合中的任一种情形出发，都推出矛盾。具体而言，肯普分下面 4 种情形证明。

- ❑ 若有一个国家与 2 个国家相邻，则可导出矛盾。
- ❑ 若有一个国家与 3 个国家相邻，则可导出矛盾。
- ❑ 若有一个国家与 4 个国家相邻，则可导出矛盾。
- ❑ 若有一个国家与 5 个国家相邻，则可导出矛盾。

只要证明了上述情形，那么因为任何正规地图中都避免不了会出现上述构形之一，于是任何情况下都会有矛盾。

从情形一与情形二导出矛盾是比较容易的事情，下面我们以情形二为例。

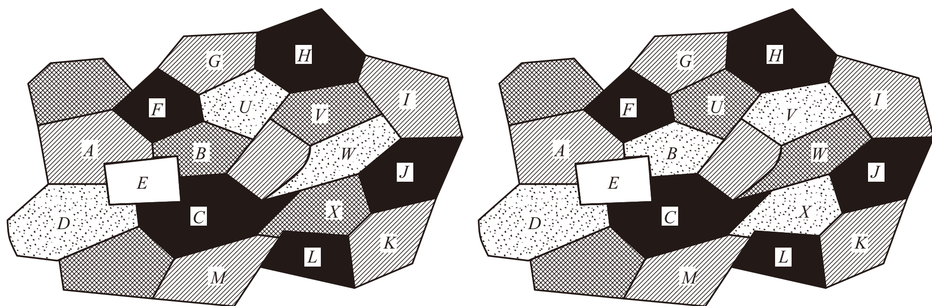
如下面的左图所示，在一幅最小正规五色地图上，有一个国家 D 与 3 个国家 A、B、C 相邻。我们把 D、C 合并成一个国家 C'，如下面的右图所示。



最小正规五色地图表明的是再减少一个国家就能用 4 种颜色来染色，因此根据新地图的国家数比原图少，可得出新地图能用 4 种颜色染色。现在，把这幅染了四色的地图恢复，然后把 D 染上与 A、B、C' 不同的另一种颜色。这样，我们就得到了原图的一种四色染色方案！这与原图是最小五色地图矛盾。

再来看复杂的情形三。

如下图所示，假设有一幅最小正规五色地图，国家 E 与 4 个国家 A、B、C、D 相邻。与上面类似，把 E 与某一邻国合并起来看成一个国家，新地图便有一种四染色方法。如果 A、B、C、D 所使用的颜色少于 4 种，那么用另一种颜色染 E，便可得到原地图是一幅四色地图而非五色，从而得到矛盾。因此，剩下要考虑的只是 E 的邻国染 4 种不同颜色的情况。图中分别用、、、表示红、黑、蓝、黄 4 种颜色。



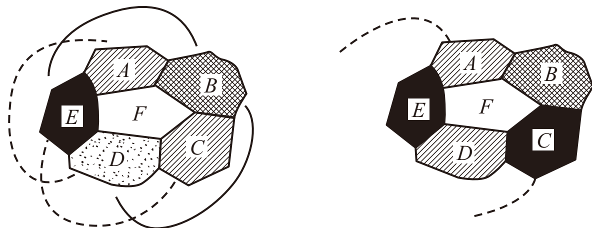
为处理这种更复杂的情形，肯普引入了一个“肯普链”的想法，即考察由 A 到 C 的一连串红黑相间的相邻国家链或由 B 到 D 的一连串蓝黄相间的相邻国家链。肯普首先证明了，E 的这两对邻国（A、C 为一对，B、D 为另一对）中，必有一对中不存在上述的相间路。否则可以推出，从 A 到 C 的红黑相间路与从 B 到 D 的蓝黄相间路上必

有一个公共国家将染上两种不同颜色，这与染色规则矛盾。于是，我们不妨设 B 到 D 不存在蓝黄相间的相邻国家链。如前面的左图所示，图中只有从 B 到 X 的蓝黄相间路 $BUVWX$，这条路没有通到 D。

然后，使用两色互换技术：把蓝黄相间路 $BUVWX$ 中的蓝黄两色调换，得到黄蓝相间路 $BUVWX$（如前面的右图所示）！调换后，B 被染成了与 D 相同的黄色，而 E 的邻国只染了三种颜色红、黑、黄，因此只需把 E 染上蓝色，便可得到此最小五色图的一个四染色方法。矛盾。

于是，我们得到情形三，即“若有一个国家与 4 个国家相邻，则可导出矛盾”。

沿着相同的思路，肯普探讨了情形四。其大致思路是，设有一个国家 F 与 A、B、C、D、E 五国相邻，与前面的分析类似，我们只需考虑合并了国家 F 后的那张四色地图里，A、B、C、D、E 分别染 4 种不同颜色的情形。不妨设它们分别染红、蓝、红、黄、黑，如下面的左图所示。肯普下一步考虑的是在 A、B、C、D、E 中寻找不含相间路的两对点。他指出，若 B 到 E 有一条蓝黑相间路，B 到 D 有一条蓝黄相间路，那么从 A 到 D 必无红黄相间路，从 C 到 E 也必无红黑相间路。于是，使用两色互换技术，把从 A 出发不到达 D 的红黄相间路中的红黄色调换，使得 A、D 同为黄色；把从 C 出发不到达 E 的红黑相间路的红黑调换，使得 C、E 同是黑色。由此，得到 F 的邻国只用了三种颜色（如下面的右图所示），把 F 染上第四种颜色，如红色，于是便得到原地图的一种四染色方法。这与最小五色地图的假设矛盾。



大功告成！四色问题得证。这就是肯普在其论文中给出的极为精巧的证明。为了

更清晰地理解，我们再简单总结一下其证明思路。

先找到一个由 4 个构形组成的不可避免的集合，然后在“最小正规五色地图”存在的假设下，从每一种构形出发都推出矛盾。为此，他反复使用了如下思路：给定一个构形，然后考虑包含这一构形的任意一个正规地图。通过合并区域，得到的新地图因区域减少能有 4 种颜色着色，然后运用技巧说明含有这一构形的原地图也能四着色。具有这种特点的构形（含有这种构形的地图可以约简成有更少国家的地图）后来被称为可约构形。显然，最小正规地图中若存在这种构形，一定会导出矛盾。因此，可约构形也可以理解为在最小正规五色地图中不会出现的一种区域安排。

于是，我们可以用最简单的术语概括肯普的工作：找到一个由可约构形组成的不可避免集，又由不可避免集中的 4 个构形都是可约的导出矛盾，从而证明四色定理。

肯普的证明得到凯莱等数学家的认可，人们认为四色猜想已经被证明了。肯普的工作还激发了其他研究者。在他发表自己的证明之后不久，1880 年英国的泰特就发表了一个以图论为基础的新证明。

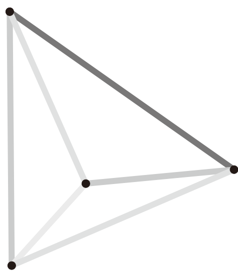
泰特的证明

泰特（1831—1901）的思考途径是把四色问题转化为与之等价的图论问题，然后借助图论中的内容完成证明。



泰特

泰特考虑的是边染色，具体来说就是用 3 色对一个 3 正则图的边染色，这种染色现在已被称为泰特染色。所谓 3 正则图，就是指一个图中的每一个顶点都是 3 次的（即从每一个顶点都可引出 3 条边）。下图即为一个对 3 正则图的 3 边染色的简单例子。



对于更复杂的 3 正则图，是否还能进行泰特染色呢？泰特做出一个猜想：每个 3 正则 3 连通平面图都有泰特染色，也就是说这样的图都是 3 边可染色的。

为了明白泰特猜想的意思，我们还要解释一下 3 连通。连通图的意思，我们已经介绍过了。对于一个连通图来说，连通程度可能不同。比如，有的连通图可能移去一个顶点就会成为不连通的图，而有的连通图则可能要移去多个顶点才能变为不连通的图。于是，使一个图不连通所需要移去的最小顶点数就可以衡量图的连通程度，而这

个数被称为图的连通度。连通度为 3 的图称为 3 连通图。

泰特证明了，自己提出的泰特猜想与四色猜想是等价的。因此，只要证明了泰特猜想，就可完成四色问题的证明。

为此，泰特又成功证明了一个结果：每个哈密顿的且 3 正则的图都有泰特着色。于是泰特猜想成立只需确定：每个 3 正则 3 连通平面图都是哈密顿的（即有哈密顿圈）。泰特相信这是明显正确的，由此他最终完成了四色猜想的证明。其推导过程是，“每个 3 正则 3 连通平面图都是哈密顿的”，而“每个 3 正则哈密顿图都有泰特着色”（已证），因此泰特猜想即“每个 3 正则 3 连通平面图都有泰特染色”成立。又因为泰特猜想等价于四色猜想（已证），因此四色问题解决。

推理过程的每一步似乎都无懈可击，但这一证明有一个真正的薄弱环节，即大前提“每个 3 正则 3 连通平面图都是哈密顿的”是否真的正确。这一结果在直觉上似乎成立，泰特本人正是相信这一结论正确，因而深信自己已经证明了四色猜想。然而，直觉上的成立并不等于证明。因为没有严格证明这个大前提，泰特的证明是不完全的。

第四节 失败与成功

1852 年，格思里提出四色猜想。随后是 20 多年的沉寂期。19 世纪 70 年代，随着凯莱正式提出这个问题，它开始广泛流传并进入一个研究高潮。在四色猜想的热门年代，很快出现了上述的肯普、泰特的证明，四色问题似已成过去式。然而，多年后风波骤起，这两个证明都被发现有错误存在。尘埃落定的四色问题再度引起人们关注。

光荣的失败者

1879 年，肯普在其 8 页论文中，给出了四色猜想非常巧妙的证明。在它发表后，数学界公认肯普已经圆满地解答了四色问题，甚至一所中学的校长把这个问题列为校内数学竞赛的有奖征解题，并附加条件，要求答案长度“不超过一页，30 行算式，另可多加一页图”。

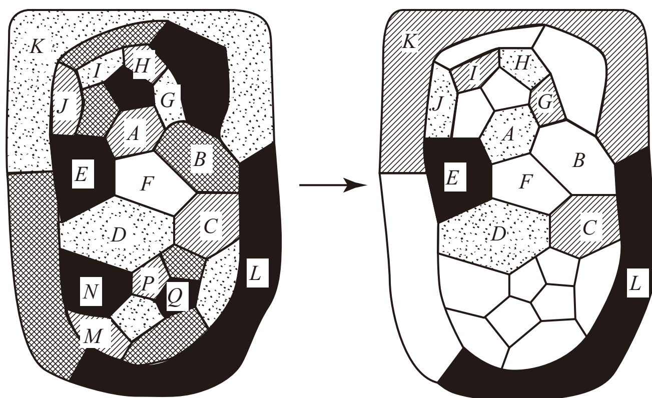
然而，1890 年，肯普证明发表 11 年后，一位英国年轻人的 7 页论文在数学界引起轰动。在这篇题为“地图染色定理”的论文中，作者希伍德

声称肯普的证明是有漏洞的！希伍德发现的这一漏洞出在肯普证明的最后一步上。

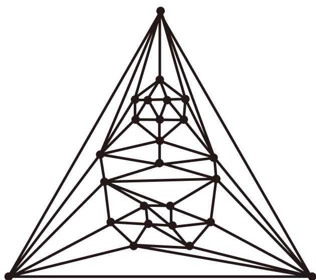
在上一节中，我们提到在情形四中，肯普认为，若 B 到 E 有一条蓝黑相间路，B 到 D 有一条蓝黄相间路，那么从 A 到 D 必无红黄相间路，从 C 到 E 也必无红黑相间路。于是，他把从 A 出发不到达 D 的红黄相间路中的红黄色调换，使得 A、D 同为黄色；把从 C 出发不到达 E 的红黑相间路的红黑调换，使得 C、E 同是黑色。

问题出在，当“把从 A 出发不到达 D 的红黄相间路中的红黄色调换，使得 A、D 同为黄色”的过程中，这里有可能会出现一条从 C 到达 E 的红黑相间路！在这种情况下，再调换 C、E 的颜色，得到的 F 邻国仍是 4 种颜色，也就无法进一步推出矛盾。

为了容易理解，我们先给出一个局部反例。如下面的左图所示，F 有 5 个染有 4 种颜色的邻国，并且其中 A 到 D 不存在红黄相间路，C 到 E 也不存在红黑相间路。按肯普的做法，把由 A 出发不到达 D 的红黄相间路 $AGHIJK$ 的红、黄两色调换，得到下面右图所示的黄红相间路 $AGHIJK$。然而，问题出现了，在这一调换后，原本 C 到 E 之间不存在红黑相间路的情况发生了改变，在 C、E 之间已经出现了一条红黑相间路，即图中所示的 $CLKE$。这时，再调换 C、E 的颜色，F 的邻国仍为四色。



希伍德本人给出的反例图如下，这个有 25 个顶点的反例图成为图论最著名的例图之一。

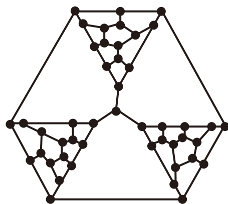


有必要指出，希伍德的反例，并不是肯普所证明的命题（四色定理）的反例，而是肯普所给的证明的反例。也就是说，希伍德的反例图仅仅指出肯普证明的错误，并不是说这个图不能用四种颜色着色。后来有些数学爱好者误解了这个反例的含义，把证明这个反例图能四着色当作重大课题，并且认为证明这个例图能够四着色就能推翻希伍德的结论。几年前国内就出现过典型的一例。2003 年,《科技日报》报道了一位毕其一生精力志在解决“四色猜想”的小人物董德周的故事。文章作者被这位小人物追求理想、献身科学的执着精神深深感动。然而，非常遗憾的是，董先生做出的突破是没有多大价值的。事实上，董先生所做的仅仅是用简单的方法“推翻了希伍德的反例，证明了该反例能四着色而非五着色”。董先生指出“该反例能四着色”是正确的。然而，他错在这一正确的结果并没有“推翻希伍德的反例”，更不意味着“四色猜想的彻底解决”。

希伍德用一个反例指出肯普证明中的重大缺陷，肯普接受了这一指正，并在伦敦数学学会上亲自报告了希伍德的工作，同时他称自己无法弥补这一漏洞。其他人也无法修补这一微妙但致命的错误。因为无法清除漏洞，肯普的证明被否定了。虽然如此，数学家们仍然赞扬“肯普的论证是极为聪明的，虽然他的证明是不完全的”。我们后面会看到，正是沿着肯普的基本思路，数学家们在为修补其漏洞做了不懈努力后，最终促使了 100 年后的成功。

肯普的证明在 11 年后被否定了。泰特的证明又如何呢？他那需要填补的大前提的证明后来找到了吗？

1946年，也就是在泰特发表其证明66年后，加拿大数学家塔特（1917—2004）的一篇论文对其做了终审判决。判决结果又一次出乎人们的意料：这个大前提的证明是不存在的，因为它根本不正确。理由非常简单，塔特找到了一个3正则3连通的非哈密顿图！可以检验，下面这个有46个顶点的图是3正则3连通的，但可以证明它不含哈密顿圈。塔特最初给出的证明很复杂，后来在1972年，一位中学教师给出了此图不含哈密顿圈的一种简洁灵巧的证明。塔特所发现的这个在数学史上有重要地位的图，后被称为塔特图。



塔特的结果完全动摇了泰特证明的根基，它表明沿泰特的途径无法证明四色猜想。这一否定结果是一项极为杰出的工作，数学界谈论了好久。

虽然泰特的证明是错误的，但在证明中，他提出的一些有趣的想法引出了一些新的研究领域。人们对图论中边染色的兴趣正是由他的研究所激发的。此外，泰特的尝试给数学家另一有益的启示。19世纪以后，不少数学家试图用证明四色猜想等价命题的方法来攻克这一问题，他们证明了很多与四色猜想等价的猜想。1931年，数学家惠特尼证明了四色猜想成立等价于每一个哈密顿平面图的顶点可4着色。1943年，数学家哈德维格尔提出猜想，每一个连通的 n 色图可以收缩到完全图 K_n，当 $n=5$ 时，这一猜想等价于四色猜想。

一部数学史，是失败与成功交织的历史。在四色猜想证明的征途中，肯普与泰特以其“有漏洞的伟大证明”而成为光荣的失败者。

希伍德的贡献

在四色问题的历史上，继肯普、泰特之后，接下来的一个比较重要的人物是英国数学家希伍德（1861—1955）。希伍德曾在牛津大学学习，他因其大胡子以及允许他的狗参加其讲座而出名。令人印象更为深刻的是他对四色问题的投入。自 1890 年发表第一篇有关四色问题的论文起，在其生命的后 60 年中他一直研究它，并先后发表 7 篇这方面的重要论文。他一直工作到 78 岁才退休，在 85 岁时还向伦敦数学学会提交他最后一篇关于这个问题的研究论文。



希伍德

在 1890 年的论文中，希伍德指出了肯普的错误，否定了肯普关于四色猜想的证明。作为一种补救，希伍德运用肯普的技巧证明了，每张平面地图着色只用 5 种颜色就足够了，这被称为五色定理。

五色定理

下面我们换一种角度给出五色定理的另一种说明。

考虑正规地图的对偶图，于是正规地图的五色定理可转化为其对偶图中与之等价的命题：对于任何平面图 G 上的点而言，我们都能用 5 种颜色来着色，使得相邻的两点不同色。

为了证明此命题，我们使用一种直接证法：数学归纳法。

显然，若平面图 G 的顶点数 $n \leqslant 5$ ，命题显然成立。

假设顶点数为 $n-1$ 的平面图可用 5 种颜色染色，根据数学归纳法，下面我们只需说明顶点数为 n 时命题同样成立即可。

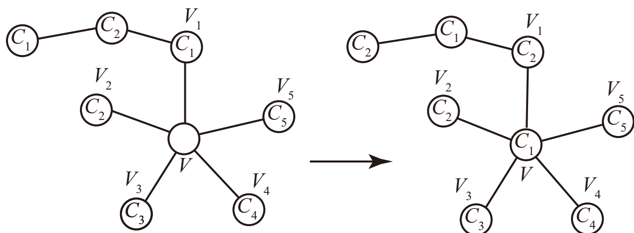
在正规地图中，我们有结论“至少包含一个边数少于 6 的国家”，这一结论转换为对偶图中等价的结果，则有“平面图中一定存在一个顶点 V，与其关联的边数少于 6”。

现在在具有 n 个顶点的平面图 G 中去掉顶点 V 及与 V 关联的边，得到一个顶点数只有 $n-1$ 个顶点的新图。由归纳法的假设知，新图必可用 5 种颜色染色。然后要考虑加入 V 点以后的情况。

显然，如果与 V 关联的边小于 5，则 V 的相邻点用的颜色数小于 5，于是只要把 V 染上与相邻点不同的颜色，问题即可解决。因此，需要考虑的是与 V 关联的边恰为 5 的情况。在这种情况中，如果 V 的邻点颜色有相同的，那么 V 的相邻点用的颜色数小于 5，同样只要把 V 染上与相邻点不同的颜色，问题也已解决。这样，只剩下一种最复杂的情况需要考虑：与顶点 V 关联的 5 个顶点 V_1、V_2、V_3、V_4、V_5 已经染了 5 种不同

的颜色 C_1、C_2、C_3、C_4、C_5，此时，无法直接对 V 染色。

借用肯普的技巧，考虑由 V_1 通达 V_2 的 $C_1C_2C_1C_2C_1C_2\cdots$ 相间路。如果没有这样的相间路，那么只需运用两色互换技术，于是，V_1 与 V_2 可同时染上 C_2，余出的 C_1 可为 V 点染色。问题解决。



如果存在由 V_1 通达 V_2 的 $C_1C_2C_1C_2C_1C_2\cdots$ 回路，又如何处理呢?

显然，若在 V_1、V_2、V_3、V_4、V_5 任两点之间都存在一条两色的相间路，则意味着这 5 个顶点之间形成了一个 K_5 或 K_5 的剖分图。由前面我们简单提到过的波兰数学家库拉托夫斯基的结论可知，G 包含了一个 K_5 的剖分图，因此不是平面图，这与 G 是平面图矛盾。因此，5 个点之间不可能任两点之间都存在一条两色的相间路。于是，我们就可以对那没有两色相间路的两个点运用两色互换技术，从而解决问题。

根据数学归纳法，我们得出五色定理成立。

希伍德染色定理

通过上面的介绍，我们看到希伍德在对四色问题的研究中，既指出肯普的错误，又退一步证明了五色定理。此外，他还把注意力转到其他曲面上图的色数确定问题，并提出了新的问题：四色猜想是一个平面图的染色问题，如果把这幅图画在曲面上，对地图作相应的染色，会有什么结论呢？

对于球面染色而言，人们早已知道它等价于平面染色。当把问题转到更复杂的曲面上时，将涉及刻画曲面的一个重要拓扑不变量：亏格。比如，平面或球面的亏格是 0，环面的亏格是 1。亏格为 h 的曲面从拓扑角度来说等价于一个有 h 个洞的曲面，记为 S_h。

希伍德着手研究的正是 S_h 中图的染色问题。他把在这种曲面上染色所需的最小色数记为 $\chi(S_h)$，然后他试图求出 $\chi(S_h)$ 的表达式。如果能做到这一点，作为特殊情形的四色猜想就可作为推论得出来。

结果，他在 1890 年的论文中成功推出证明。

$h>0$ 时，$\chi(S_h)\leqslant H(h)$，其中，$H(h)$ 表示 $\left[\dfrac{7+\sqrt{1+48h}}{2}\right]$，$[x]$ 表示不大于 x 的最大整数。显然，他获得了 $\chi(S_h)$（h 为正整数时）的一个上界。非常遗憾，根据希伍德的证明，这一公式对 $h=0$ 不适用，因此这一漂亮结果恰恰在处理众所瞩目的四色问题上派不上用场。否则，结合人们早就知道的平面染色不能少于 4 种颜色，就可直接推得四色猜想成立了。

在得到上述公式后，希伍德做出一个猜想：$\chi(S_h)=H(h)$，$h\geqslant 0$。

如果这一猜想能够证实，代入 $h=0$ 即可得到 $\chi(S_0)=4$，即四色猜想成立。然而如

上所提到的，因为希伍德得到的有关 $\chi(S_h)$ 的上界公式对 $h=0$ 不适用，因此沿这条途径无法解决 $h=0$（即四色猜想）问题。

在 $h>0$ 方面，则只需证明 $\chi(S_h)\geqslant H(h)$，再结合希伍德已证明的结果，即可推出 $h>0$ 时，希伍德猜想成立。不难理解，为了证明 $\chi(S_h)\geqslant H(h)$，只需要在曲面 S_h（$h>0$）上找到一个地图 M，使 $\chi(M)=H(h)$ 即可。

在这方面，希伍德取得了一些进展。他发现，在环面上可以画出 7 个国家，其中任何一个国家都与其他 6 个国家相邻。按照染色原则，每一个国家的颜色，都不能和相邻国家的颜色相同。因此，这个图必须用 7 种颜色染色（如下面的左图所示，先对环面沿两个方向剖开后得到的矩形 7 着色，中图是把矩形卷成一个圆筒形，最后把圆筒的上下两端合拢，得到下面最右侧的图所示的环面 7 着色图）。这意味着，对亏格为 1 的环面来说，$\chi(S_1)\geqslant 7$。



另外，由希伍德的上界公式可知，对亏格为 1 的环面有：

$$\chi(S_1)\leqslant H(1)=\left[\frac{7+\sqrt{1+48\times 1}}{2}\right]=7$$

两者结合起来，可得到环面上有七色定理成立，即任何环面地图要染色，至多只要用 7 种颜色就够了。

沿着这条途径，到 1968 年时，人们证明了：$\chi(S_h)=H(h)$，$h\geqslant 2$。

由此得到希伍德的猜想 $\chi(S_h)=H(h)$，在 $h>0$ 时成立，这一结论被称为希伍德地图染色定理。一个唯一的例外是亏格为 1 的克莱因瓶。对于这一著名的曲面，希伍德公式给出的结果是 7，但数学家富兰克林证明了克莱因瓶上的每个地图染色都只需用 6 种颜色。

于是，到 1968 年时，只剩下 $h=0$ 的情况未加证明。这真是数学中的怪现象。按一般人的观念，更复杂的曲面（亏格大于 0）应当更困难，可是当 $h>0$ 的所有情形都已解决时，对于最简单的曲面（即平面），人们却束手无策。

更复杂、更高阶曲面都圆满解决了。唯独看起来最简单、最低级、最初始的四色问题，没有获得圆满解决。而那些解决高阶曲面获得成功的好方法，都无法对付这个最简单、最低级、最初始的四色问题。这真令人深感惊奇与困惑！著名数学大师陈省身在谈到这个问题时曾不断地说，这是“惊人的”“费解的”“神秘的”。

着迷于四色问题的希伍德将他一生 60 余年的光阴全都花费在这个问题的研究上。如上所介绍的，他把平面上的地图染色问题推广到更复杂的曲面地图染色问题，并得出一些结论。然而，在四色问题上他没有再取得更多进展。事实上，在肯普的证明被否定后的几十年中，这个看似简单的小问题难倒了许多著名数学家，许多解决这一难题的尝试都以失败而告终。

一个著名的故事发生在德国一流数学家闵可夫斯基（1864—1909）身上。这位真正的数学神童在年仅 17 岁时，就曾完成了一篇 140 页的论文，并在第二年赢得了极负盛名的法国科学院大奖。1902 年的一天，一位学生在课堂上向这位素以谦虚著称的老师问起四色问题。正上拓扑课的闵可夫斯基回答说：“四色问题之所以还没有解决，是因为世界上一流数学家都没有花工夫去研究它。”兴之所至，他当堂就证明起来。他写了几黑板，仍然没证出。下一节课，他继续推导，仍然“挂了黑板”。就这样，一连几个星期过去了。他开始感到自己已经身陷泥沼。最后，在一个阴雨的早晨，闵可夫斯基走进教室时，忽然雷声大作。闵可夫斯基站在讲台上，面朝着学生，自嘲地说：“看来，老天也在责备我狂妄自大，我也没法解决这个问题。我们还是返回来上我们的拓扑课吧。”接着，他就从数周前中断的地方开始继续讲起拓扑学。

第五节
四色足够

地图着色是人人都容易明白的。“给平面地图着色，使得有公共边界线的区域着不同颜色，四种颜色足够”，这一问题简单、清楚、具体。但这个看来毫不复杂的问题又有其惊人的困难。自肯普的证明和泰特的证明被发现漏洞后，四色问题重新成为一道向人类智力提出挑战的世界难题。

放电理论

1913 年，美国数学家伯克霍夫（1884—1944）在四色问题研究中迈出一步。在一篇有关图的可约性的文章中，他改进了肯普的约化方法，在肯普的基础上引进了一些新技巧，证明比肯普构形更大的一些构形是可约的。1920 年，富兰克林应用伯克霍夫的想法，证明当国家数目不超过 25 时，四色猜想成立。1926 年，温恩把国家数提高到 27。1940 年，温恩再将其提高到 35。1970 年，国家数被提高到 39。1975 年，国家数被提高到 52。1976 年，这一纪录刷新为 95。花费半个多世纪，可以用四种颜色着色的地

图的国家数才仅仅提高了几十个。这种推进确实是太缓慢了。国家数增加得很慢，原因是国家每增加一两个，不同国家之间边界关系的类型就变得复杂得多，而证明的关键就是必须把地图的所有类型都考虑到，不能有遗漏。

更大的问题是，沿这种研究途径，即便国家数得到很大的提升，仍然无法从根本上解决四色问题，因为由此只获得量的积累，无法得到质的突破。换言之，这种向无穷缓慢推进的处理方法，将永远不可能填补 95 个区域的地图与任何能想到的由无限多个区域组成的地图之间的空缺。

如何才能彻底解决四色问题呢？肯普曾给出一条有效的途径，寻找不可避免的可约构形集合。只不过，在处理“一个区域恰好被 5 个区域环绕”这一构形时，肯普失败了。他证明这种构形是可约的，但希伍德指出他的证明有缺陷。因此，肯普给出的不可避免构形集合中并非所有构形都可约，而一个不可约构形的存在宣告了他的证明失败。也就是说，肯普指出一条可行的路，但自己并没有走通。在其他途径走不通的情况下，部分数学家重新折回到肯普的方向上，他们仔细分析了肯普受挫的最后一种情形，认识到必须寻找与肯普所找不同的不可避免可约构形集。如果能找到一个不可避免的可约构形集合，那么四色问题就能够获得证明。不过在实际寻找的过程中却遇到了意想不到的计算困难。经过几十年的徘徊不前后，德国数学家希斯（1906—1995）最先打开了新的局面。

希斯在 1936 年就开始钻研四色问题，1950 年，他公开宣称可以通过寻找可约构形的一个不可避免集来解决四色问题。为了寻找不可避免的可约构形集合，希斯第一次提出了一种具体可行的算法：“放电”方法。下面我们简略介绍一下。

首先，把平面正规地图染色的研究转化为其对偶图——三角剖分中顶点染色的研究。然后，他把三角剖分看成一个电网，对每一个顶点赋予电荷。因为顶点数是 2、3、4 的情形肯普已经解决了，因此只需要考虑那些次数为 5 以及高于 5 的顶点。若对每个 d 次顶点赋以 $6-d$ 个电荷，则只有 5 次顶点得到正电荷，而次数大于 6 的顶点都被赋以

负电荷。可以很容易证明：任何三角剖分图中所赋的电荷数之和总为正数。

于是在每个顶点赋以电荷后，可按照一定的规则向邻近的顶点移动电荷。在这种移动过程中，某些 5 次顶点可能会失去一些正电荷，即被放电；而某些大于 6 次的顶点可能会得到一些正电荷，即被充电。但无论如何移动，整个网络中带电总量不会改变，即在移动后，电荷数之和仍为正数。这表明网络中总会有一些带正电荷的顶点。于是，我们可以根据放电的规定，在不知道整个地图的情况下，分析并确定出那些带正电荷顶点邻域的局部布局，从而最终找出所有只带正电的构形所组成的一个构形集合。因为每个网络都有正电荷的顶点，而这个构形集合中已考虑了所有带正电顶点的构形，因此我们找到的这个构形集合是一个不可避免构形集。

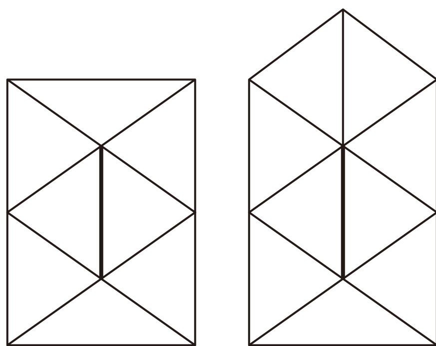
举一个简单的例子说明一下。把 $\frac{1}{5}$ 单位电荷从每一正电荷顶点（5 次顶点）移到它的每个负电荷邻点（即每个 7 次或 7 次以上的顶点）上去，看看会出现什么情况。

首先，一个 8 次或更高次数的顶点，即便它的全部邻点都是 5 次顶点，也不可能变为正的。这可以通过简单的计算说明：设顶点的次数是 d，它从每个 5 次顶点接受 $\frac{1}{5}$ 个正电荷，于是它能接受的最大正电荷数是 $\frac{1}{5}d$。因为它自身的电荷数是 $6-d$，所以放电过程结束时，它所带的电荷至多是 $(6-d)+\frac{1}{5}d=\frac{30-4d}{5}$，由 $d\geqslant 8$，可知这个电荷数总是负的。也就是说，在这一放电规定下，8 次或更高次数的顶点最后不可能带正电荷。

而且，一个 6 次顶点永远不会变成带正电荷的，因为它的初始电荷为 0，绝不会接受正电荷。因此，只需要考虑 5 次顶点与 7 次顶点。

先看 5 次顶点的情况。在这个过程中，如果一个 5 次顶点的邻点都带负电荷，那么它要向外至少移动 1 个正电荷，这样它本身最后已不可能再带正电荷。因此，一个 5 次顶点在这一放电过程后仍保持正电荷，仅当它的邻点至少有一个次数不大于 6。这会出现两种情况：邻点有一个 5 次顶点，以及邻点有一个 6 次邻点。这意味着，在这

种放电过程中，只能出现两种构形，如下面的左图所示，5 次顶点与另一个 5 次顶点相邻；或如下面的右图所示，5 次顶点与一个 6 次顶点相邻。



最后，再看 7 次顶点的情况。对于一个 7 次顶点来说，仅当它至少有 6 个 5 次邻点时，最后才带正电荷，因为 $6\times\frac{1}{5}+(-1)=\frac{1}{5}>0$。而此时，其中至少有两个 5 次顶点是相邻的，因此会出现上面左图所示的情况。

通过以上分析，我们看到，“把 $\frac{1}{5}$ 单位电荷从每一正电荷顶点（5 次顶点）移到它的每个负电荷邻点”这个简单放电过程产生的不可避免集由上图所示两个构形组成，其中一个由一条边连到另一个 5 次顶点上，另一个由一条边连到一个 6 次顶点上（如图中粗线所示）。因为上面的论证对任意网络都成立，所以在任何地图中都可以找到这两个构形中至少一个，这也正是不可避免集表示的意思。

通过这一个简单的例子，我们可以得知：因为每幅地图中总有正电荷，所以无论放电规定如何，网络中总会有带正电荷的点，而选择出现在正电荷邻域的每种类型中的构形，就可以构成不可避免集。当然，给出的放电规则越复杂，得到的不可避免构形的集合也就越复杂。事实上，希斯当时估计他要寻找的不可避免集大约包含 10 000 个构形！

寻找不可避免集还只是完成问题的第一步。下一步是要判断集中的构形是否可约。为此，希斯设计了一个计算机程序，用来验证构形的可约性。但是，这种验证需要计

算机计算相当长的时间，每个大约需要 100 小时，而在他的估计中要验证的构形约为 10 000 个。这意味着，用计算机也得花上 100 年左右。前景依然渺茫。而且，只要在这不可避免集中发现有一个构形是不可约的，证明就是无效的，整个计算将全无价值。肯普的证明中出现过这样的情况。20 世纪 70 年代左右曾有类似的情况发生。当时，一位数学家构造了一个构形，如果这一构形是可约的，四色定理即可获证。于是，在这一构形是否可约尚未做出完全判定之时，四色问题已被解决的消息就被传开了。可当将这一工作交给当时运算速度最快的计算机进行验证后，得出一个遗憾的结果，这个构形不是可约的。于是，寻找四色问题解决方法的尝试又一次失败了。

虽然多次努力仍未成功，但到 20 世纪 60 年代末时，人们在这一问题上已经获得一些进展。一方面，希斯的思想为解决四色问题开辟了一条可行之路。另一方面，人们认识到解决四色问题可能要借助计算机。这两种思想的结合在不久后就结出了硕果。

四种颜色足够了！

第二次世界大战后，希斯在讨论班上公布自己的方法时，一名美国大学生哈肯恰好参加了这期讨论班。被四色问题吸引的哈肯很快投入到四色问题的研究中。

1970 年，哈肯发现可以改进希斯的放电方法。1972 年 5 月，哈肯在伊利诺斯大学（他自 1962 年就受聘于该校）一个讨论会上谈到解决四色问题非靠计算机不可，他还谈到使用计算机可能遇到的各种困难，他又表示虽然知道困难所在，但不知道怎样才能克服这些困难。在座的有一位在计算机程序方面功底非常深厚的名叫阿佩尔的人，

他认为计算机可以解决这个问题。此后，两人开始了他们的合作。后来，他们还得到一名计算机科学专业研究生的帮助，这名研究生为他们编写了有效的程序来验证可约性。哈肯、阿佩尔采用的新的计算机实验方法在第一轮就获得了许多有用信息。随后他们修改程序以克服缺陷并反复试验。6 个月后，他们开始确信，他们的方法是行之有效的。1975 年，他们设计的程序从搜索状态开始向目标发起最后的冲击。1976 年 6 月，在用三台电脑花费了 1200 个机时后，他们终于成功地构造了含有 1936 个（后被削减为 1476 个）可约构形的不可避免集。

哈肯、阿佩尔在多伦多大学举行的 1976 年的美国数学学会和美国数学协会的夏季会议上，向全世界宣布了他们的成功。后来，他们又提交了一份报告，刊登在《美国数学学会通报》9 月刊上。1977 年，两人的论文《每个平面地图都可以四着色》在美国《伊利诺斯数学杂志》上发表，全文长 139 页，还附上电子计算机程序的微型胶片 400 页。当其成果发表时，从美国伊利诺伊州寄出的信上都特地盖了纪念邮戳“Four colours suffice”（四种颜色就够了）。

四色猜想被证明了！这条数学家们会以香槟酒干杯祝贺的消息，很快传遍了全世界，引起了轰动。一个历时 120 多年的世界难题终于被攻克了！

证明的余波

四色问题在提出一个多世纪后终于获得解决了，然而其解决方式建立在对电子计算机的使用与依赖上。在此之前，计算机主要作为数学家的辅助工具而被使用。与之

不同，哈肯、阿佩尔的证明中很大部分且关键部分是由计算机完成的，其中用到的一些概念就是计算机证明的产物。于是，四色猜想成为第一个主要由计算机完成证明的重要数学难题。这一引人注目之处在数学界引起了巨大争议。

按照一般程序，一个数学证明在正式发表之前，编辑要找有相当水平的数学家进行评审。而在其发表后，任何有充分耐心的数学同行都能对这个证明进行核对并证实它是正确的。然而，哈肯、阿佩尔的计算机证明却无法使用通常的审查方法，因为它无法向人们展示出符合传统检测要求的证明过程。事实上，在其论文正式发表之前，编辑部采用的方法是，将哈肯和阿佩尔的程序输入一台独立计算机，通过运行得出相同的结果来证明其论文的正确性。这种非正规的审查过程激怒了一些数学家，他们认为这是一种不可信的核查方法。要知道，每台现代计算机的软件和硬件中都有隐匿的缺陷，而且每台计算机还有可能出现瞬间即逝的失误。因此，当计算机做了上千小时的计算时，谁能保证不存在由于计算机的某种突然不规则运行而产生的逻辑错误呢?另外，一个长达 400 页的计算机程序又如何保证其正确性呢？事实上，1981 年，施密特在其博士论文中，将哈肯、阿佩尔证明中的计算机程序的 40% 进行了检验，确实发现了 14 个小错误和一个较大的错误。当这篇博士论文传送到世界各地时，还引发了“计算机没有证明四色问题”的传闻。当然，问题没有这么严重。事实上，那些小错误可以在几分钟改正，较大的错误也只需要几天就能改正。面对许多人的疑问，1986 年，哈肯、阿佩尔两人写了一篇文章“四色问题的证明是充分的”并将其刊登在《数学信使》杂志第 8 卷第 1 期上。文中试图介绍这一证明的原始思想，让读者了解怎样得到这一证明，并解释为什么在检验影响细节时突然出现的错误不会影响证明的正确性。他们最后表示:“我们将对证明中剩下的 60% 加以独立的核实，而且欢迎有人找到进一步错误时通知我们……当这些完成后，我们打算出版原始证明的一种完全校订本。”

面对“这一无法按传统意义核对的计算机证明是否有效”的怀疑与批评，接受这

一证明的数学家指出，“无论如何，从实用的观点看……与大量事件在人工校核中出现错误的机会相比，计算机出错的概率是极其微小的”。而对四色定理计算机证明有效性的最后一点疑问后来也消除了。1993年，罗伯森、托马斯等数学家把构形简约为633个，从而给出四色定理的简化证明。2004年，数学家沃纳和冈席尔用通常的数学证明验证程序，对罗伯森等人的证明进行了验证，肯定了其证明的正确性。

虽然随着时间的推移与计算机在数学中影响力的加大，哈肯、阿佩尔的证明被数学界普遍接受了，但这一项惊人的成就自始至终未能赢得多数数学家的赞赏。因为这种“蛮力式”的机器证明在数学家们看来缺乏洞察力，不如传统的高雅证明那样明晰与发人深省，失去了数学的美感。如中国著名数学家苏步青所说，机器证明“即便是真的，我们总觉得没有什么数学味”。确实，在数学家们看来，一个数学证明不仅回答了问题，它还使人们对为什么答案应该如此有所理解。把问题送进一个黑匣子，然后从另一端收到一个答案，这增加了知识但没有增进理解力。数学家真正喜欢的是一个阐明性的定理。他们期待着某些令人眼前一亮的新洞见、一种核心深处的想法之美、使人耳目一新的证明。然而，哈肯与阿佩尔的回答是“把这个问题分解成数千种情形，然后将所有情形一个接一个地在计算机上运行，这样就完成了证明”。对于哈肯、阿佩尔的证明，有人这样评论：“一个好的数学证明应当像一首诗，而这纯粹是一本电话簿。”因此，许多数学家感到沮丧与失望是非常自然的。更符合数学家口味的足够优美和简明的人工证明会在某一天被发现吗？有这种可能，但也可能根本不存在。也许数学中就存在单靠人脑解决不了而必须用计算机才能解决的复杂问题。在这种情况下，就出现了一类新的有意义的、无法用传统形式证明的定理。从这个观点看，四色定理呈现出另一种价值，即它帮助人们看清了单靠人脑解决问题的有限性。

随着四色定理计算机证明逐渐获得数学界的认可，它引发的另一个争论是，在什么程度上人们可以说一个依赖于大量的、非人力所能控制的计算机证明真算一个证明呢？换句话说，接受四色定理的计算机证明，传统意义上的“证明”概念是否也发生

了改变呢？在此之前，人们始终把证明看作是一种逻辑上严密可靠的推理过程。借助这样的推理，数学家就可以使其他人相信某个判断的真实性。一个数学家在看懂一个证明以后，就会对所论命题的真实性深信不疑，同时也就豁然理解确保这种真实性的推理。事实上，一个证明之所以成为证明，恰恰是因为它提供了这样一些可被数学同行所验证的有说服力的推理。在许多人看来，计算机证明的出现意味着整个“数学证明”的概念发生了变化：一直作为辅助作用的计算机从数学家手中接过了从事真正数学证明的重任。但在持不同观点的数学家看来，数学证明中引入计算机“这一事实并没有改变数学证明的基本概念。改变的不是数学理论，而是数学实践”。

于是我们看到，四色问题作为第一个人机合作完成的著名数学问题，在数学界和哲学界都引起了广泛关注，什么是数学证明等重大问题也因此纷纷提出并得到深入讨论。目前对这些问题作出一个回答为时尚早，也许时间会告诉我们：四色定理计算机证明的意义，不仅仅是解决了一个历时百年多的难题，更重要的是成为数学思想发展史上一系列新想法的起点。

机器证明与吴方法

四色定理的计算机证明，还推动了一门新兴学科——机器证明的发展。正如哈肯与阿佩尔所说：“我们相信，一定存在很有数学意义而只能用计算机方法来证明的定理，即便四色定理不是这样一个问题，它也说明了证明这种定理可能需要做些什么。”四色定理机器证明发表 12 年后，1988 年 12 月，加拿大林永康教授领导的小组宣称他

们用一超级电脑，用整整三年数千小时机时成功解决了一个世纪难题：证明了不存在10 阶的有限射影平面。机器证明又一次显示了它的威力。然而，这样的证明只是对某一个具体数学问题的证明，方法仅仅适用于这一具体的问题，因此只能称为“特例机证”。另一条或许更有意义的机器证明途径是，寻求用计算机求解一类问题的通用方法，即寻找“类证”——对一类定理提供一种统一的算法，使得该类定理中的每个定理都可依此方法给出证明。

1977 年，在四色定理被证明之后一年，我国著名数学家吴文俊院士（1919—2017）在《初等几何判定问题与机械化证明》的论文中，首次公布了用计算机证明数学定理的“吴方法”，从而在几何定理机械化证明方面首次取得重大突破。下面我们简单介绍一下他的思想。



吴文俊

吴文俊迈出的第一步是几何代数化。具体而言，借助解析几何的手段，把几何问题转化为代数问题，将几何定理的叙述用代数方程的形式重新表达，于是证明问题就转化为判定假设的代数方程是否能推出结论的代数方程问题。用吴文俊本人的话说，就是“从事几何定理证明时，首先取适当的坐标，于是几何定理的假设与终结通常都

成为多项式方程，称之为假设方程与终结方程。满足定理假设的几何图像，就相当于假设方程组的一个解答或零点。要证明定理成立，就是要证明假设方程的零点也使终结多项式为零”。不过，几何问题转化为纯代数问题后，并不见得简单，更不能说就能实现机械化。计算机技术的发展只是为实现数学定理证明的机械化提供了某种前提，而沿几何代数化这条路线，真正实现机器证明，还需要数学家设计出切实可行的、行之有效的“好的”计算机算法。

吴文俊真正的突破正来自于此。在 1977 年发表于《中国科学》杂志的论文中，他提出了一个新的代数方法。这个后来被国际上称为“吴方法”的新方法在证明等式型几何定理时效率比以前的方法高得多。1984 年，吴文俊又出版专著《几何定理机器证明的基本原理（初等几何部分）》，明确建立了各类几何的机械化定理，系统阐明了几何定理机械化证明代数方法的基本原理，奠定了数学机械化的基础。其开创性工作，被认为是机器证明的里程碑式贡献。2001 年，吴文俊获首届中国国家最高科技奖，而机器证明是获奖的主要原因。2006 年，由于“对数学机械化这一新兴交叉学科的贡献”，吴文俊又获“邵逸夫数学科学奖”，这一国际性大奖是国际数学界对他“数学机械化研究”的承认与肯定。

使用吴方法，可以在计算机上迅速证明一些很不简单的几何定理。1989 年，我国访美学者周咸青在其出版的专著中列出了用吴方法证明的 512 个几何定理。除了可证明初等几何定理外，吴方法还能证明局部微分几何定理，并有其他众多的应用。

吴方法的问世，使计算机自动推理研究领域发生了一场革命，不但使具有数千年历史的初等几何研究真正跨入机械化阶段，而且使几何定理机器证明的代数方法研究空前活跃起来。10 多年后，多种机器证明的代数方法提出并得到成功实现。20 世纪 90 年代，张景中院士等人又成功地解决了计算机可读性证明问题，即由计算机生成便于人们理解、掌握和检验的证明。这项工作被自动推理领域的同行誉为“计算机处理几何问题发展道路上的里程碑”“自动推理领域 30 年来最重要的工作”。不久，杨路教

授又成功地解决了几何不等式的机器证明问题。这样，初等几何定理的机器证明问题基本上由我国数学家解决了。

在计算机自动推理的研究中，几何定理机器证明曾经是最不成功的领域，现在却成为最成功的领域。这一历史的转折，缘于吴方法的出现。吴方法开辟了数学机械化的一个新纪元，但这方面的研究目前仍处于起步阶段。在更广阔的领域中实现机器证明，使数学研究进一步机械化，是摆在数学家面前的新课题。可以说，由吴文俊所倡导的“数学机械化和机械化数学”在国际范围内正蓬勃发展。

结　　语

在本章中，我们简要介绍了四色问题由提出到解决的历史。当问题画上句号之时，我们可以简单回顾一下人们的研究历程了。

1852 年，弗兰西斯·格思里提出了陈述简单的四色问题。1878 年，英国数学家凯莱向伦敦数学学会正式提出这个问题，其后四色猜想成了世界数学界关注的问题。1878~1880 年，肯普和泰特两人分别提交了证明四色猜想的论文，宣布证明了四色定理，人们认为四色猜想解决了。然而，在多年后两人的证明都被指出有漏洞。人们逐渐认识到，这个貌似简单的问题，其实是一个可与费马大猜想相媲美的难题，横亘在数学家面前的不是一块小石头，而是一座大山。当更复杂、更高阶曲面上的着色问题都被圆满解决后，平面四色问题这个平凡却又深藏陷阱的迷人问题吸引了越来越多的研究者。

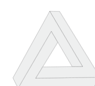
被这一问题所吸引的数学家（也包括一些世界上一流的数学家）纷纷参与到四色问题的大会战中。为了解决这一难题，数学家们花费了大量的时间，经历了多次的失败，虽然在最初一个多世纪中四色问题缺少实质性的进展，但数学家们的付出没有白费，其收获还是颇丰的。收获之一是，在对四色问题的研究中，人们找到了数十种四色问题的变体、等价说法和各种推广。由此开辟的一条证明四色问题的道路是，证明任何一个与四色问题等价的猜想。可惜沿这条新路，没有人能打开缺口。更重要的是，研究四色猜想的意义逐渐超越了解决这一数学难题本身。正是在对四色猜想的探索中，数学家们顽强前进，从一片蛮荒之地开拓出一条条道路。在这一过程中，很多种方法被提出，很多数学技巧被发展出来。这些有不少成了数学新思想的起点，并由此产生了新的理论。事实上，对四色问题的研究，大大促进了图论的发展（比如图染色研究就是由四色问题激发的），催化了组合拓扑学的诞生，而且对有限射影几何等也产生了积极的推动作用。这些都是四色问题为数学孵出的金蛋。

但在四色问题的研究上，数学家却陷入山重水复的困境。终于，希斯的出现，使人们看到一条可以翻越这座数学大山的可能道路。沿着这条道路，最终哈肯、阿佩尔获得了成功。我们可以看到，最后的成功正是建立在前人成果基础上的。顺利解决整个问题的基本思路是寻找一个可约的不可避免集，而它来自肯普的有漏洞的证明。正如美国数学家阿佩尔和哈肯所说："我们的方法基本上是肯普证明中正确部分的推广，这部分在过去 100 年中一直受到数学家们的密切注意。"在肯普之后，希斯的放电理论是沿着肯普的思路迈出的又一关键，它是四色猜想能成功用于机器证明的关键，是证明四色猜想的核心要素之所在。1976 年，哈肯、阿佩尔在改进放电理论并借助计算机后，最终获得了成功。四色问题由猜想变成了定理。然而，这一依赖计算机完成的独特证明却在数学界引发了巨大争议与讨论。在这一证明已被数学界普遍接受的今天，一个更进一步的问题是，四色定理的计算机证明是否会对数学发展带来深刻的影响，它会成为新的数学思想的起点吗？时间会告诉我们答案。如果答案是肯定的，那么我们可以说这将是四色问题为数学孵出的更大金蛋。

1993年，罗伯森、托马斯等数学家给出四色定理的简化证明。但是，新的证明中仍然包含不能人工核查的计算机步骤。于是，四色定理的证明提供了数学中极为平常的古怪事件的又一个例子：容易理解的问题与惊人复杂的证明的结合。这真是奇妙又困扰人的事。

那么，“平面地图只需要4种颜色着色”，这个能被学校孩子理解的问题，能否找到更好的方式说明其成立的理由呢？或者说，是否存在不依赖计算机的简洁的人工证明呢？

当四色猜想以意外的方式解决，并变成了四色定理后，众多数学家退出了研究的行列。然而，许许多多为这一极易理解的问题所吸引的数学爱好者仍在苦苦寻觅那或许存在或许不存在的人工证明。可以说，它在数学爱好者中的流行程度堪比几何三大问题。以至于图论专家哈拉里对此评论说：“四色猜想真可以改名叫‘四色病’了。因为它像传染病一样，很容易传染。有时是良性的，有时却是恶性的或慢性的。还没有发明一种预防针可对付这种病，但一个人的体格如果足够强壮，那么稍一感染就可以终身免疫。这种病会反复发作，虽然还没有致死的记录，但它会使人痛苦非凡。我们已经观察到，这种病至少一次从父亲传染给了儿子，所以也许是会遗传的。”当然，对四色定理人工证明的寻求与用尺规完成几何三大问题是有差别的。对于后者，我们可以断言其绝无可能。但对于前者，我们却只能持谨慎的存疑态度。也许吧，一种非常巧妙的不借助计算机的四色问题证明会在某一天问世。但对于投身于这一研究中的爱好者，我们有两点建议要提。一点是，“体格不够强壮”的不要去尝试医治这种“四色病”。另一点是，在做自己的研究之前，了解前人的成果是有益的。然而可惜的是，事实上众多研究者往往是置已有的研究于不顾，结果因不了解四色问题难解的症结所在，掉入陷阱或误入歧途却不自知。



第五章

费马问题

第一节
从毕达哥拉斯到丢番图

人类对自然数的认识曾经历了一段极其漫长的时间。随着自然数的出现，两类性质截然不同的问题摆在人们面前。一类是数的应用，即计算问题，也就是运用加减乘除运算得出结果，以解决实际问题。另一类是数的理论问题，也就是研究数的性质及数与数之间的特殊关系。前者与实际问题紧密相关，受到古代多数民族的重视并获得了发展。但对后者的研究，却主要是古希腊人的贡献，他们由研究个别抽象的数过渡到研究任何可能的数，从而产生出一个称为数论的独立分支。

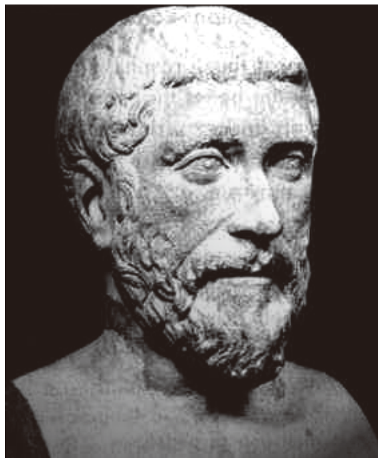
这个独特的数学分支中有着大量有趣又极难解决的问题，我们在本章中要介绍的费马最后问题就是其中著名的一个。为了了解它的来龙去脉，让我们追溯到数论的源头。

毕达哥拉斯与毕达哥拉斯学派

沿着“数的理论问题”这一方向迈出最初几步的是古希腊的毕达哥拉

斯及其领导的毕达哥拉斯学派。对这个数学史中会反复提到的著名数学家及其学派，我们有必要简单介绍一下。

毕达哥拉斯（约公元前580—前500），出生于希腊一个小岛上。他早年曾师从阿那克西曼德。后曾游历过古埃及、古巴比伦等东方国家，在这些国家他不断向有学问的人请教，以丰富自己的见解和知识。当他完成游历回到自己的故乡时，已是一个思想成熟的智者了。但在家乡他无法施展才志，于是他移居到意大利半岛南部的克罗托内，在那里他赢得了人们的信任与景仰，并组织起一个政治、宗教、数学合一的秘密团体，即后来所称的毕达哥拉斯学派。这个团体后来在政治斗争中遭到破坏，毕达哥拉斯逃亡到塔兰托，后被杀害。



毕达哥拉斯

毕达哥拉斯是古希腊哲学家、数学家、天文学家，还是音乐家、教育家。他的思想与学说对希腊文化产生的影响是巨大、深远和多方面的。后人尊称他为“智慧之神”。

毕达哥拉斯一手创建的毕达哥拉斯学派具有深远的历史意义。在他的领导下，该学派进行了多方面的研究工作。在数学方面，其影响更为巨大。有许许多多的研究都

是从他们开始的，他们极大地推进了数学的发展。可以说，毕达哥拉斯是数学这门学科的奠基人、创始者。最有名的是以他名字命名的毕达哥拉斯定理（即我们所说的勾股定理）。通过毕达哥拉斯定理而发现无理数，是这一学派在数学上的另一重大贡献。

开辟对数的性质的研究也是毕达哥拉斯及其学派的众多数学贡献之一。在他们的努力下，数不再仅仅用来记账和计算，数本身的价值受到了重视，并由此产生了古希腊所称的“算术”，即我们现在所说的“数论”。顺便指出，这正是后面我们将会提到的丢番图、高斯的数论名著分别叫作《算术》《算术研究》的原因。

在这种研究中，他们提出了许多独特而有趣的数论问题，奠定了古希腊数论的雏形。比如，他们引入并研究了三角形数、正方形数等与图形相关的所谓“形数”；他们对自然数的可除性问题进行了深入的研究。通过研究，他们把自然数划分为若干类，如我们不陌生的奇数与偶数的划分，以及素数与合数的划分。此外，他们还做出过更为奇特的划分，并从中发现了一些有趣的数，如亲和数、完全数等。

此外，他们还对与勾股定理密切相关的勾股数进行了研究。我们所熟悉的勾股定理是，在一个直角三角形中，斜边的平方等于两直角边的平方之和。设 x、y 为两直角边的长，z 为斜边的长，那么这一定理可用数学符号表示为 $x^2+y^2=z^2$。而所谓勾股数（或称毕达哥拉斯三元组）是指三个恰好满足这一不定方程的正整数组合，比如 (3, 4, 5)、(5, 12, 13) 等。如何找到更多的三元组？毕达哥拉斯学派对此进行研究时，感兴趣的不是给出个别的勾股数，而是想找到产生许多勾股数的公式。他们获得了部分成功，他们发现如果 m 是一个奇自然数，则 $\left(\frac{m^2+1}{2}\right)^2=\left(\frac{m^2-1}{2}\right)^2+m^2$，因此 $\left(m,\ \frac{m^2-1}{2},\ \frac{m^2+1}{2}\right)$ 可以给出勾股数。如果奇数 $m=2n+1$，我们可以得到一个等价的形式，即 $(2n^2+2n+1)^2=(2n^2+2n)^2+(2n+1)^2$。通过后面这一形式可以看出，这一公式只能产生斜边与一个直角边差是 1 的勾股数解。

约公元前 380 年，古希腊著名哲学家柏拉图设计出另一公式：$(m^2+1)^2=(m^2-1)^2+$

$(2m)^2$，其中 m 为任意的自然数。这一公式的特点是斜边与长直角边的差为 2，但它也不能给出所有的勾股数。那么，是否有一种公式能产生全部勾股数呢？

可以证明，设 m 和 n 是两个正整数，m 大于 n，m、n 互素且奇偶性不同，那么 m^2+n^2、m^2-n^2、$2mn$ 就可以给出全部基本勾股数。至于基本勾股数，是指三个数互素。显然，在基本勾股数的基础上，同乘以一个相同的倍数，就可以得出所有勾股数了。

在我国，能产生全部基本勾股数的法则最早出现在《九章算术》一书“勾股”一章中。在印度，数学家婆罗摩笈多在公元 628 年左右写成的重要著作《婆罗摩修正体系》中，也明确给出了上述勾股数公式。

而在西方，最先掌握这一法则的数学家是古希腊的丢番图。

丢番图与数论

在第一章中，我们已经介绍过丢番图及他在代数学方面的贡献。他在代数学发展上做出的贡献使他赢得了“代数学之父”的称号。除此之外，丢番图在数论方面取得的突出成就也对后代产生了巨大影响。事实上，他划时代的《算术》一书虽包含了代数内容，具有代数学的特征，但更是一本数论著作。这书除了第一卷外，其余的问题几乎都是考虑未知数比方程数还多的问题，我们把这种问题叫不定方程。为了纪念丢番图，现在通常把这类方程叫丢番图方程，求解不定方程则被称为丢番图分析，而由求解不定方程发展起来的当代两大先进手段也被称为丢番图逼近和丢番图几何。

勾股方程可谓最古老的不定方程，对于这一不定方程的全部解法，丢番图是了解的，在他的名著《算术》中有若干题就用到了，只是没有明显表达出来。

在体现作者天才的《算术》一书中，丢番图给出了部分数论命题，如形为 $8n+7$ 的整数不能表示为三个平方数之和等。但他更多的是研究诸如“求两个平方数，使其和是一个立方数”（第四卷第 3 题）之类的问题。他感兴趣的是正有理数解。这与现在的丢番图方程研究稍有差别，现在一般只考虑整数解。

作为一位有高超技巧的解题能手，丢番图在多方面显示出惊人的睿智和独创性。他关于不定方程的一些巧妙解答，往往使人眼花缭乱、叹为观止。

然而，他生活于希腊数学的晚期，伟大的古希腊数学在他之后不久就落下了帷幕，随之西方数学步入停滞不前甚至倒退的黑暗时代。因此，他的思想很快湮没了，很长的时期里都没有对数学产生大的影响，甚至他的巨著也成了残本。然而经过漫长的 1000 多年的等待后，他的巨著终于在 15 世纪被重新发掘。韦达在其基础上把代数学推进了一大步，而另外一些数学家则从中看到了美妙的数论。于是，丢番图及其《算术》成为许多大数学家研究数论的出发点，并为这些数学家的洞察力和灵感提供了无尽的源泉。

第二节
从费马到高斯

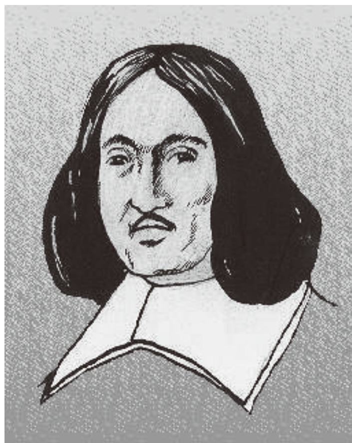
毕达哥拉斯及其学派埋下了西方数论的种子。随后欧几里得在《几何原本》中发展了它（在后面的章节中我们将会介绍），再之后是天才的丢番图以及他的巨著《算术》问世。然而，其后西方数学（包括数论）陷入停滞不前的状态。1000 多年后，当丢番图的六卷残存本出现在他身后知音费马的书桌上时，一切都改变了。

出谜者：业余数学家之王费马

费马，1601 年 8 月出生在法国图卢兹一个富有的皮革商人家中，逝世于 1665 年 1 月。

费马最初的职业是律师，1631 年被任命为图卢兹议会的顾问——请愿者接待室的一名顾问。如果本地人有任何事情要呈请国王，他们必须首先使费马或他的一名助手相信他们的请求的重要性。作为顾问的费马提供了本省和巴黎之间极重要的联系。除了在本地和国王之间起联络作用之外，

他担任的这一职务还要保证发自首都的国王命令得以在本地区执行。费马另外的职务包括在司法部门的工作，后来他还当过图卢兹最高法院的大法官。资深的他足以处理最困难的案件。根据各种流传的说法，费马是一位称职的文职官员。



费马

令费马本人深感幸运的是，他担任的职务使他可以孤立于图卢兹社交界，从而为他提供足够的闲暇时间，因为任此职务的人要回避社交活动，以防止腐败。于是，这个个性淡泊的人度过了没有任何传奇经历的平静一生，并在业余时间专心致志于自己爱好的数学，最终谱写出了数学史上最美妙的故事之一。

费马在年近三十时才开始认真研究数学，并且只是利用业余时间从事这种研究。然而这并不妨碍他在数学上取得累累硕果。他在几何学、概率论、微积分和数论等众多数学领域都留下了自己的足迹。

和笛卡儿同时或比笛卡儿早，费马得到了解析几何的要旨，因而与笛卡儿分享着创立解析几何的荣誉；他与帕斯卡在一段有趣的通信中一起奠定了古典概率论的基础，因而与帕斯卡被公认为概率论的创始人；他提出光学的“费马原理”，给后来变分法的研究以极大的启示；他是创建微积分学的杰出先驱者。

任何人，即便只是完成了上述工作中的某一项，就足以在数学史上留下不朽名声，更不用说能同时拥有这众多成果了。然而，费马的成就尚不止于此，他将更多的时间与精力奉献给了自己最喜爱的数论。

费马对数论的兴趣正是由丢番图的名著激发起来的。1621 年，巴谢校订注释的《算术》希腊－拉丁文对照译本在法国出版。费马买了一本，并被其中的数论问题深深吸引住了。这本书成了他的课本，值得庆幸的是，这版书留有宽大的页边空白，于是这书同时成了费马的笔记本。在研究丢番图的问题和解答时，他会受到激励去思考和解决一些其他相关的、更微妙的问题，然后他会草草在这些页边空白上以评注的方式写下自己的研究心得。


DIOPHANTI
ALEXANDRINI
ARITHMETICORVM
LIBRI SEX,
ET DE NVMERIS MVLTANGVLIS
LIBER VNVS.
CVM COMMENTARIIS C. G. BACHETI V. C.
& obseruationibus D. P. de FERMAT Senatoris Tolosani.
Accessit Doctrinæ Analyticæ inuentum nouum, collectum
ex varijs eiusdem D. de FERMAT Epistolis.

TOLOSÆ,
Excudebat BERNARDVS BOSC, è Regione Collegij Societatis Iesu.
M. DC. LXX.

让现代人有些难以理解的是，费马的数学成果在其生前几乎都没有公开发表。事实上，他只是把自己的研究写在书的页边上，或者把自己的成果按照当时流行的风气，以书信的形式向一些有学问的朋友报告。原因在于，费马是被“发现新的数学奥秘”

这类强烈的念头所驱使的。研究数学于他而言是一种消遣，纯粹是出于他对数学的爱好。当他自己能够得到新的未被他人触及的结果时，他会获得真正的愉悦与自我满足。因而公开发表和被人们承认对他而言没有多大意义。有意思的是，这位缄默的天才有时喜好捉弄人，他经常在信中向其他数学家发出挑战，要求他们证明自己发现的某些数学结果。有一次费马向英国数学家们提出过一个问题：证明方程 $y^2+2=x^3$ 仅有一组正整数解 (3, 5)。关于这个问题，他在一个注记中说，要找有理分数解是不难的，但是他已经发现一个新颖的方法，有了它，他能求该问题的正整数解。

于是，当费马去世时，他的研究成果散落在旧纸堆里，或书页的空白处，或给朋友的书信中，面临着失传的危险。幸亏他的长子意识到父亲业余爱好所具有的重要意义，决心不让世界失去父亲的发现。他先是花了 5 年时间收集他父亲的注记和信件，检查他那本《算术》书的页边空白处草草写下的字迹。最后他将这些注记在《算术》的一种特殊版本中发表。1670 年，他出版了《附有 P. 德・费马的评注的丢番图的算术》，其中包括费马所做的 48 个评注。1679 年，他又整理出版了费马的第二卷著作，在这一卷中包括抛物形求面积法、极大极小及重心的论述、各类问题的解答，这些内容后来成为微积分的一部分。费马的研究成果终于在他去世后得以流传。这些研究成果极大地丰富了 17 世纪的数学宝库，直接推动了后来数学的发展。

在他的所有贡献中，数论方面的工作最能显示出他的才华并且对后人影响最大。在他之前，数论大体上是一些问题的汇集。是费马振兴了数论的研究，提出了为数可观的数论定理，并在数论中引入一般的方法和普遍的原理，从而把数论引上了近代的轨道。可以说，正是他系统化数论后，数论才真正开始成为一门数学分支。费马也因奠定近代数论的基础，而被当之无愧地称为“近代数论之父”。作为数论发展史上一个承先启后的人物，在高斯名著《算术研究》出版之前，数论的发展始终是跟费马的推动联系在一起的。如数学史家 E. T. 贝尔在《数学精英》中所评价的：“费马是一个第一流的数学家、一个无可指摘的诚实的人、一个历史上无与伦比的算术学家。”

然而，不幸的是，费马往往写下一些必要的东西证明他已明白问题的解法，然后就不再费神写出证明的剩余部分。于是，在后人看到的注记中，他对自己提出的结果或者根本没有任何解释，或者仅仅给出对背后证明的一点点提示。这意味着，在数学界重新发现他的证明之前，他的每一个评注只能被当作猜想。而补全所有细节就作为一种挑战留给后来的数学家。接受这一挑战的是另一位数学巨人：欧拉。

数学家之英雄：欧拉

欧拉（1707—1783），18 世纪数学界的中心人物。在整个数学史上，他也是屈指可数的顶尖数学家之一。人们一般把他与阿基米德、牛顿、高斯并列为数学史上最伟大的 4 位数学家。数学家纽曼在 1956 年称他是“数学家之英雄”。



欧拉

1707 年，欧拉出生在瑞士的巴塞尔。他的父亲曾希望他学习神学，因此欧拉 13 岁时考入巴塞尔大学神学院。但欧拉的兴趣广泛，对哲学、天文学、诗歌等都有所涉猎，在数学方面更是很早就表现出过人的天赋。在名师指点下，19 岁的欧拉便开始发表高质量的数学论文并获得了法国科学院的奖金。

1727 年，欧拉成为俄国圣彼得堡科学院的成员。1741 年，欧拉应腓特烈大帝的邀请，成为柏林科学院院士。俄国叶卡捷琳娜二世在位期间，欧拉应邀重新返回圣彼得堡。后来他一直住在俄国。1783 年 9 月 18 日，欧拉与朋友们吃饭。那天天王星刚发现不久，欧拉写出计算天王星轨道的要领，还和自己的孙子逗笑。喝完茶后，突然疾病发作，烟斗从他手中落下，他口里喃喃自语“我死了……”，就这样停止了呼吸，终止了计算和生命，终年 76 岁。

欧拉有着超人般的记忆力，他不但能够记住前 100 个素数，而且还能记住所有这些素数的平方、立方，甚至四次方、五次方和六次方。他还可以进行复杂的心算。如法国物理学家阿拉戈说的：“欧拉计算时似乎毫不费力，就像人在呼吸，或鹰在翱翔一样轻松。”在完全失明后，欧拉的记忆力对他后期的研究起到了重要作用。

除过人的天赋外，欧拉还有坚强的意志力。18 世纪 30 年代中期，欧拉的右眼开始失明。1771 年，欧拉的另一只眼也失明了。当数学界认为欧拉将要悲惨地结束自己的数学研究生涯时，欧拉却仍然不屈不挠地进行着自己的研究，向他的助手口授他奇妙的方程和公式，在助手的帮助下，继续从事数学著述。正如失聪没有阻止贝多芬的音乐创作一样，失明也同样没有阻止欧拉的数学探索。在漫长的数学研究中，欧拉的遗产是无与伦比的。他的著作数量极多，产出速度极快，甚至在他完全失明以后也是如此。他博大精深和空前丰富的著述令人叹为观止。

据统计，这位迄今为止最多产的数学大师（另一位在多产方面可与他相提并论的著名数学家是埃尔德什，我们将在下一章中提到）一生发表论文共计 856 篇，专著 31 部。在这众多著作中，既有难度很高的专著，也有写给初级者的读物。非常值得一提

的是，欧拉所有著作的论述都非常清楚、易懂，并且他所选用的数学符号，都是为了将他的意思表达得更加清晰明了，因而许多由欧拉引入的数学符号成为数学中的通用符号，例如 π（1736 年）、虚数单位 i（1777 年）、表示自然对数底的 e（1748 年）、sin 和 cos（1748 年）、tan（1753 年）、Δx（1755 年）、表示求和的符号 ∑（1755 年）、用 $f(x)$ 表示函数（1734 年）等。

在欧拉极为广泛的涉猎中，数论也是其极为重视并做出辉煌贡献的领域。

欧拉最初对数论并没有太留意。他对数论的兴趣是由哥德巴赫引发的。哥德巴赫，这位现在以哥德巴赫猜想闻名的人物早年学习法学，后来在大学中攻读医学和数学，1710 年游历欧洲，结识数学家莱布尼茨、伯努利等人，1725 年移居俄国，任彼得堡科学院会议秘书兼数学教授。1727 年，他到莫斯科当上沙皇彼得二世的家庭老师，1742 年起担任外交部公使。自 1729 年起，他与欧拉进行了长期的书信交往。哥德巴赫曾被数论问题深深吸引，尤其是费马提出的那些美妙的却没有证明的结论。不过，他的才能却不足以赶上他的热情，于是在与欧拉的通信中，他把有关费马未证明的猜想告诉欧拉。不久后，欧拉就被数论，特别是被费马一系列未证明的猜想深深迷住了，他开始尝试证明费马那些精妙的评注。事实上，在欧拉有关数论的著作中，有相当一部分旨在证明费马的猜想。

在欧拉的努力下，数论又有了进一步发展。其后，数学家拉格朗日与勒让德又都沿着欧拉的工作路线前进，在数论方面做出了贡献。勒让德还在 1798 年出版了第一部数论教科书——《数论》，后经多次修订补充，完善于 1830 年。但在数论研究中迈出更大步子的则是德国的伟大数学家：高斯。

数学之王：高斯

高斯（1777—1855），数学史上最伟大的数学家之一。他有着非凡的数学天赋，有两个有关他的广为流传的故事可以证明这一点。



高斯

据说，高斯不满3岁时，有一天，他父亲念念叨叨地结算他管辖的几个工人的周薪，突然，小高斯打断他说，爸爸，你算错了，应该是……老高斯再用纸和笔核对了一下，发现自己的确错了，孩子的答案是对的。而那时谁也没教过高斯算术。另一个故事更为人所熟知。在他7岁的时候，老师让学生从1开始一直加到100。当其他孩子还在费力地一个数一个数相加时，高斯已很快在他的小石板上写出答案，放到了老师的大讲堂上。最后，老师检查学生的结果，发现大多数学生的答案都是错的，而交卷最快的高斯的答案却是准确的。

后来领地的公爵听说自己领地上出了这么一个聪明好学的孩子，就资助高斯进

一步学习。于是，不满 15 岁的高斯得以进入卡罗琳学院学习。1795 年 10 月，18 岁的高斯离开故乡到哥廷根大学学习。1796 年 3 月 30 日，他用圆规与直尺作出了正十七边形。这一成功为他赢得了巨大的荣誉，同时也确定了他一生的研究方向。在此之前，高斯曾在研究数学还是研究语言学方面徘徊过。这一巨大成功使他非常兴奋，决定把数学作为自己终生的事业。甚至他还希望在自己的墓碑上刻上一个正十七边形，以纪念他年轻时的这项重要发现。高斯的希望因各种原因未能实现，但在他的家乡不伦瑞克，为他竖立的纪念碑上的的确确刻上了正十七边形，算是圆了他的遗愿。而 1989 年第 30 届国际数学竞赛在哥廷根大学举行时，也特意设计了正十七边形的会徽。

1799 年，高斯以大学毕业论文证明了数学中的代数基本定理，并因此赢得了“代数学之父”的称号。1801 年，高斯出版了数学经典著作《算术研究》。

1855 年 2 月 23 日，高斯在哥廷根天文台的住所逝世。在他逝世后，汉诺威王命令为他铸一个纪念奖章。汉诺威著名的雕刻家和奖章制作者布雷·默尔做成了一枚 70 毫米的奖章，上面刻着：汉诺威王乔治 V. 献给数学之王。

自那以后，高斯便以“数学之王”著称。对后人来说，这位被誉为“能从九霄云外的高度按某种观点掌握星空和深奥数学的天才”，其无与伦比的数学天赋往往令人深感敬畏。美国数学家 G. F. 塞蒙斯曾评价说：“这就是高斯，一个至高无上的数学家，他在那么多方面的成就超过一个普通的天才人物所能达到的水平，以致我们有时会产生一种离奇的感觉，以为他是‘上界天人’。”

在高斯一生涉及的众多研究领域中，数论是他的最爱。正如他那众所周知的名言所说：“数学是科学的皇后，而算术[①]是数学的皇后。”在他看来，数学在科学中占有特别的重要地位，而数论又在数学中占有特别重要的地位。高斯对数论的贡献体现在他

① 数论在早期称为算术。到 20 世纪初，才开始使用数论的名称。

最伟大的专著《算术研究》中。在这本书中，高斯把数论中的记号标准化，把当时存在的定理系统化并加以推广，把尚待研究的问题以及已知的解题方法进行了分类，还引进了一些新方法，如同余理论。另外，他关于正多边形的作图方法也写在这本书中。此后 100 多年中，几乎所有数论方面的发展都能追溯到这部著作所引出的思想。可以说，这部著作是数论领域的划时代巨著，它为数论的研究迎来了一个新纪元，奠定了现代数论的基石。

第三节
最深奥的数学之谜

在上一节中我们看到，欧拉对费马的论断一一进行研究并加以证明，拉格朗日和勒让德都沿着欧拉的工作路线前进，而高斯则使数论成为统一理论。然而，无论是欧拉、拉格朗日、勒让德，还是数学之王高斯都没有揭开由费马留下的一个最深奥的数学之谜。

数学史上最撩人的页边评注

前面我们已经介绍到，以研究数论为消遣的费马在购得一本丢番图的《算术》后，进行了深入研究，并把自己的心得以评注的形式写在书的页边空白处。1670 年，费马去世 5 年后，他的儿子将这些页边评注编进一个新的《算术》版本中出版。几十年后，主要是在欧拉的努力下，费马的这些评注一个接一个地被解决了，除了一个例外。

约在 1637 年，费马读到了《算术》第 2 卷第八命题“将一个平方数分为两个平方数”，灵感迸发的他想到了一个更一般的问题。于是，他在这题

目的旁边空白处用拉丁文写下了一段具有历史意义的文字："将一个立方数分为两个立方数，一个四次幂分为两个四次幂，或者总的来说，将一个高于二次的幂分为两个同次的幂，这是不可能的。"用现代符号表示这段文字，即不定方程 $x^n + y^n = z^n$（$n > 2$）没有正整数解，或没有 $xyz \neq 0$ 的整数解。这就是费马问题，又叫费马猜想。我国较普遍地称为费马大定理，以与费马的另一结论费马小定理区分。而在国外更多地称为费马最后定理（Fermat's Last Theorem），或简记为 FLT。之所以被称为"最后"定理，是因为它是需要被证明的费马评注中的最后一个。

一个可以叙述得如此简单和清晰以至于小学生就可理解的问题，后来却成了巍然屹立在数学家面前的一座珠穆朗玛峰！为这一问题增添了格外光彩的是，紧接上面一段评注文字之后，费马写下的一句话："关于此，我确信已发现一种美妙的证法，可惜这里空白的地方太小，写不下。"人们也曾试图寻觅他的"美妙的证法"，但找遍一切可能的地方都没有找到。

于是，费马写下的数学史上最撩人的页边评注为世人留下了两个不解之谜。第一个谜是，费马猜想成立吗？这个谜在此后的三百多年间，困惑、吸引、难倒了数不尽的数学家，包括"数学家之英雄"欧拉、"数学之王"高斯等在内的一流数学家都在其面前败下阵来。直到 1995 年，这个比哥德巴赫猜想更悠久、更有名的难题才被出身英国剑桥的数学家安德鲁·怀尔斯攻克了。在后面的章节中，我们将介绍人们揭开第一个难解之谜的征程。这里我们先来探讨一下费马留下的第二个谜：他的美妙证法真的存在吗？

正如贝尔所评价的，费马"是一个第一流的数学家、一个无可指摘的诚实的人、一个历史上无与伦比的算术学家"。

确实，费马的严谨忠实的性格和他作为一个算术学家的无比的洞察力，使许多人坚定地相信：当费马宣称他可以证明他的定理时，他知道他在说什么。事实上，当他宣称有了一个证法，后来也就找到了一个证法，差不多他的所有明确断言都是如此。

而且这些断言的最终证明并非一件易事。即使如欧拉那样的天才，证明费马的某些结论也是一项很艰难的工作。比如，费马在1659年给朋友的信中声称自己证明了有关素数的一个美妙结果：每个形如$4n+1$的素数都是两个平方数之和。为了证明这一点，欧拉断断续续苦思了7年，才在1749年得到证明的方法。

当然，大师也会有失误的时候。一个典型也许是唯一的例子是费马数猜想。1640年，费马对形如$2^{2^n}+1$的数进行了研究。他发现当n取0、1、2、3、4时，对应值分别为3、5、17、257、65 537，这几个数都是素数。于是，费马在给朋友的一封信中写道："我已经发现形如$2^{2^n}+1$的数永远为素数。很久以前，我就向分析学家们指出这个结论是正确的。"费马同时坦白承认，他自己未能找到一个完全的证明。对费马所研究的$2^{2^n}+1$这种形式的数，后来人们称为费马数，并用F_n表示。费马当时的猜测相当于，所有费马数都一定是素数。1729年12月1日，哥德巴赫在写给欧拉的一封信中问道："费马认为所有形如$2^{2^n}+1$的数都是素数，你知道这个问题吗？他说他没能作出证明。据我所知，也没有其他任何人对这个问题作出过证明。"这个问题吸引了欧拉。1732年，年仅25岁的欧拉在费马死后67年得出$F_5=641\times 6\ 700\ 417$。这一结果意味着$F_5$是一个合数，因此费马的猜想是错的。更为不幸的是，研究的进展表明费马不但是错的，而且非常可能是大错特错了。此后有更多的费马数被人进行了研究。随着电子计算机的发展，计算机成为数学家研究费马数的有力工具。但即使如此，在所知的费马数中，竟然没有再添加一个费马素数。迄今为止，除了被费马本人所证实的那5个费马素数外，没有再发现一个！因此，人们开始猜想，在所有的费马数中，除了前5个是素数外，其他的都是合数。如果这一结论被证实，那么对于费马的草率猜想来说，恐怕不会有更为糟糕的结局了。顺便提一下，费马数在数学中得到深入广泛的研究与引人注意并非仅由费马的失误所造成的戏剧性效果引起，更重要的原因来自伟大数学家高斯的一项杰出发现。我们已提到，高斯在19岁时利用尺规作出了正十七边形，迈出了自欧几里得以来正多边形尺规作图方面的重要一步。在1801年的《算术研究》一书中，高斯得出更进一步的漂亮结果。他指出，一个正素数多边形，只有当是

费马素数时，才能用尺规作图。也就是说，就目前而言，可用尺规作图完成的正素数边形只有 3、5、17、257、65 537。就这样，正多边形作图问题与费马数极其密切地联结在一起了！全然无关的领域竟能以出人意料的方式彼此联系，这正是数学的一大魅力之所在。

在对费马数的猜测上，费马这位伟大的数论天才过分看重自己的直觉，轻率地做出了他的错误猜测。但如果想到，费马当时曾承认自己并未证明出自己猜想这一点，那么费马的这次失误反而更好地证实了他可靠的人品。

此外，费马确实提出过一个解决数论问题的美妙方法，即费马无穷递降法。这一方法在确立否定结果时显得特别有用。这一方法的应用可以概括如下。

为了证明不存在正整数 $a, b, c, \cdots$ 满足关系 $R(a, b, c, \cdots)$，假定相反情形成立。在此假定下，证明 $R(a_1, b_1, c_1, \cdots)$ 成立，其中 a_1 是正整数且 $a_1 < a$。接着，可以用同样方式证明 $R(a_2, b_2, c_2, \cdots)$ 成立，其中 a_2 是正整数且 $a_2 < a_1$，如此下去以至无穷。但只有有限个正整数小于 a，这是不可能的。于是引出矛盾，从而得到结论：关系 $R(a, b, c, \cdots)$ 不能为正整数 $a, b, c, \cdots$ 所满足。

为了说明费马的无穷递降法，我们考虑把它用于一个简单问题：证明 $\sqrt{2}$ 是无理数。

假定相反情形成立，即 $\sqrt{2}=\frac{b}{a}$，其中 a, b 为正整数。

由于 $\sqrt{2}+1=\frac{1}{\sqrt{2}-1}$，因此 $\frac{b}{a}+1=\frac{1}{\frac{b}{a}-1}=\frac{a}{b-a}$，于是有 $\sqrt{2}=\frac{b}{a}=\frac{a}{b-a}-1=\frac{2a-b}{b-a}=\frac{b_1}{a_1}$。但 $1<\sqrt{2}<2$，即 $1<\frac{b}{a}<2$，因此 $a<b<2a$。于是有 $a_1=b-a>0$ 且 $a_1=b-a<a$。简言之，若设 $\sqrt{2}=\frac{b}{a}$，其中 a,b 为正整数。那么我们可以得到 $\sqrt{2}=\frac{b_1}{a_1}$，其中 a_1 是小于 a 的正整数。重复这一过程，我们有 $\sqrt{2}=\frac{b_2}{a_2}$，其中 a_2 是小于 a_1 的正整数。这个过程可以无限重复下去，但正整数不可能无限减小。于是我们开始的假定是错误的。因此，

$\sqrt{2}$是无理数。

资料表明，费马运用这种方法成功证明了许多数论命题。他在给朋友的信中声称自己用这种方法成功证明了，每个形如 $4n+1$ 的素数都是两个平方数之和。而我们已提到为解决这一问题，欧拉曾断断续续花费了 7 年时间。特别是，费马用自己的方法还证明了，边长为整数的直角三角形的面积不是一个平方数，即不定方程 $\begin{cases} x^2+y^2=z^2 \\ \dfrac{1}{2}xy=u^2 \end{cases}$ 没有正整数解。

后人发现，由这一结论可以推出 $x^4+y^4=z^4$ 没有正整数解。因此，人们认为费马确实证明了指数 $n=4$ 时的费马大定理。问题是，费马证明了一般情形的费马大定理了吗？

费马提出的无穷递降法在 18 世纪成为一种证明数论问题的有用技巧，欧拉、勒让德等都用过这一方法。但可惜的是，这种方法虽然对某些指数有效，但并不能证明一般情形的费马大定理。

因此，虽然总有人相信费马“美妙的证法”的存在，但更多的人（包括伟大的高斯）都投了费马的反对票。也就是说，人们普遍相信费马并没有真正证明费马大定理，或者更可能的是，他得到的“美妙的证法”正如成千上万后来人的证法一样是一个错误的证法。关键的一点是，费马生前没有想到会把这些旁注公之于世，因此即便认识到自己的想法有错，他也不一定回头再把这个旁注加以删改了。就这样，费马留在页边上的寥寥数语给后人开了一个莫大的玩笑。我们将看到，正是这一出“错误的喜剧”给数学带来了极大的推动。

第四节 两个世纪的尝试

费马问题的叙述是如此简单易懂，给人以容易证明的假象，加上费马的神秘旁注，更增添了其魅力。于是在其提出后，无数受其吸引的数学家和业余数学爱好者致力于它的证明。在本节中，我们介绍一下对这一问题的早期探讨。

小小的第一步

我们先对要证明的费马大定理做两点简化工作。

首先，在这一方程中，若其中两个数有公共因子，那么容易推出第三个数也含有这一公共因子，因此方程两边可以约去这一公共因子。

具体而言，如果 $(x, y) = d$，那么 $d \mid z$，这样令 $x = dx_1$，$y = dy_1$，$z = dz_1$，则 $x^n + y^n = z^n$ 可以简化为 $x_1^n + y_1^n = z_1^n$，其中 x_1、y_1、z_1 两两互质。后者有解，则原方程也有解。因此，我们可以假定原方程中三个未知数 x、y、z 两两互素。

其次，如果指数 n 是一个合数，即 $n=lm$（其中 l 是正整数），那么 $x^m+y^m=z^m$ 无正整数解，则 $x^n+y^n=z^n$ 也必无正整数解。因为 $x^n+y^n=z^n$ 若有正整数解，则 $x^{lm}+y^{lm}=z^{lm}$，即 $(x^l)^m+(y^l)^m=(z^l)^m$，这意味着 $x^m+y^m=z^m$ 有正整数解，与其无正整数解矛盾。

对于一个大于 2 的整数 n，当 n 是偶数时，必为 4 的倍数或为某个素数的偶数倍，当 n 是奇数时，必是一个素数 p 的倍数，因此，证明费马大定理实际上只需证明 $n=4$ 与 $n=p$（p 为素数）时没有正整数解，即证明方程 $x^4+y^4=z^4$ 和 $x^p+y^p=z^p$（p 为素数）没有正整数解就可以了。

上面已经提到，费马用无穷递降法已经证明了 $n=4$ 的情形，后来欧拉等数学家又给出新证。剩余的问题是证明当指数是奇素数时，费马大定理成立。后来历史表明，这是非常困难的，数学家要踏上的是一条漫长而艰难的征程。

在这一征程中，首先要对付的是指数为 3 的情形。就这一小步就花费了一百多年的时间。直到 1753 年 8 月 4 日，欧拉写信给哥德巴赫，声称他已证明了这一点，但在信中没有写出证明的过程。1770 年，在《代数学入门》一书中，欧拉详细叙述了其证明，这成为费马大定理素指数研究中最早公开发表的证明。其证明方法建立在无穷递降法的基础之上，但其关键之处用到了形如 $a+b\sqrt{-3}$（其中 a、b 为整数）的数，特别是这类数的唯一因子分解定理，即在整数中成立的唯一因子分解性质在引入了 $a+b\sqrt{-3}$ 这类新数后仍然成立。欧拉运气很好，因为在这一情形中唯一因子分解确实是成立的，但这仅仅是一种巧合。后来人们知道，对于形如 $a+b\sqrt{-n}$ 的数系，唯一因子分解定理并不总成立。比如，对于 $a+b\sqrt{-5}$（其中 a、b 为整数）这类数来说，我们有 $6=2\cdot3=(1+\sqrt{-5})(1-\sqrt{-5})$，唯一因子分解定理不再成立。这也意味着，欧拉对 $n=3$ 的证明方法，不能推广到证明其他素指数。

在欧拉之后，高斯又对指数为 3 的费马大定理给出一种严格证明。但令数学家们大感困窘的是，证明出指数为 3 的情形是自费马大定理提出之后一百多年里，人们取

得的唯一成就。进展之微真是令人丧气。终于，在 19 世纪初，随着一位女士登上数学舞台，数学家们得以稍稍摆脱了这一难堪的局面。这位在早期尝试解决费马大定理并取得小突破的女性叫苏菲·热尔曼。

闯入数学王国的女性：热尔曼

热尔曼（1776—1831）出生在法国一个富裕家庭。在法国大革命的动荡岁月里，由于无法出门，热尔曼就在家中尽情阅读父亲的藏书。结果，13 岁的她被一本介绍数学历史的书吸引了，特别是书中介绍到的阿基米德的故事深深打动了她。当她看到 75 岁高龄的老人因专心致志研究数学图形，而不幸死于罗马士兵之手时，她无比震惊。她相信数学中必定有着无穷的魅力。由此，她迈进了数学王国，开始潜心钻研起数学。很快，她对数学的热情达到了夜以继日、废寝忘食的程度。父母开始担心她的健康，最后强迫她晚上早睡。为了防止她夜间偷偷起来看书，晚上故意不给她烧壁炉的木柴，还拿走了她的所有外衣。然而，热尔曼探索数学的决心和渴望没有停止。为了躲开父母的约束，她把蜡烛藏在秘密的地方。到了晚上，她会用毯子把自己裹起来保暖，摸进她父亲的书房去找书。最终，她对学习的渴望战胜了她父母的固执，被感动的父母让步了。就这样，18 岁之前通过自学，热尔曼在代数、几何和微积分方面都打下了牢固的基础。为了进一步提高，她需要到其他地方去寻求知识。



热尔曼

当时，巴黎综合工科学校刚建立，有许多一流的法国数学家在那里授课。热尔曼渴望进入这所学校，但这所学校像当时其他学校一样拒收女生。失望但不放弃的热尔曼选择了一些自己感兴趣的课程，然后从朋友们那里借来听课笔记自学。她对拉格朗日的分析学课程特别着迷，并递交了一篇自己结合这门课程所做工作的论文，递交论文时她使用了当时已放弃学业的一个男生勒布朗的名字。拉格朗日很欣赏“勒布朗”的优秀工作。当从学生那里听说，勒布朗先生其实是一位年轻小姐后，深感惊讶的拉格朗日鼓励并赞扬了热尔曼的工作，而且把她介绍给许多法国数学家。

1801 年，高斯的《算术研究》出版，热尔曼认真研读了这部名著。1804 年，她用勒布朗的化名给高斯写信，希望能与高斯交流自己对这部名著所得到的一些想法。几封信后，她的许多敏锐的洞察给高斯留下了深刻印象。由此，两人进行了长时间的通信。

1806 年，法国入侵普鲁士，法国军队占领了高斯居住的城市汉诺威。听到这一消息，热尔曼马上想到了死于战争的阿基米德，她十分担心阿基米德的不幸会在高斯身上重演。于是，她拜访了恰好是她父亲朋友的法军统帅。受感动的将军派特使去高斯家执行保护高斯的命令，以确保这位数学之王的平安。当从特使那里听到保护自己的

是“热尔曼”时，高斯被这个完全陌生的名字搞糊涂了。一封及时寄到的信使他恍然大悟。在这封信中，热尔曼解释了自己的化名，并写道：“……我以前用勒布朗这个名字给你写信，承蒙你如此宽容地给我回信，我实在是不敢当……我希望我今天向您吐露的情况，不会让我失去在我使用一个借来的名字时您曾给过我的那种荣幸，并希望您能抽出几分钟的时间告诉我您现在情况怎样。”

高斯热情的回信充满了赞扬：“在晓得一向敬仰的‘勒布朗先生’竟变成‘索菲·热尔曼’时，您可以想象我的讶异和仰慕之情了。您的表现实在令人难以置信！您对一般抽象科学及神秘数字的品位确实难能可贵，令人激赏不已。这一门壮丽无比的科学所特具的迷人风采要不是有心深入挖掘，是很难领略到的。事实上，也再没有其他东西能令我如此神往和专一。这门学问的趣味和魅力绝不是我故作玄虚，它们确实充满了我的生命，就如您对它的敬慕和雅好一样。”

1816 年，由于在弹性表面振动分析方面的工作，热尔曼荣获了法国科学院大奖。然而，在那对女性充满偏见的时代，仅仅因为是一位女性，许多大门就对她关闭了。她找不到合适的职业，终生没有获得过一个学位。直到 1831 年，在高斯的推荐下，热尔曼将获得哥廷根大学的荣誉博士学位。不幸的是，在正式授予她这个学位之前，她死于与之搏斗了两年的乳腺癌，终年 55 岁。在这位潜心于学术研究的女性死亡证明书上，身份被记为“无职业未婚妇女”。

然而就是这样一位被当时社会所忽视的女性，在费马大定理的研究方面留下了自己的印记。

在费马大定理的研究中，有一种途径是把问题分为两种情形来探讨。第一种情形是，对于素数 p，当 p 不能整除 xyz（或称 p 与 xyz 互素）时，不定方程没有正整数解。第二种情形是，对于素数 p，p 能整除 xyz 时，不定方程没有正整数解。

针对第一种情形，热尔曼非常漂亮、灵巧地证明了：若对于素数 p，$2p+1$ 也是素

数，则费马大定理在第一种情形下成立。在与一些数学家通信中，她汇报了自己得到的这个一般性结论。

热尔曼探讨的这类素数后被称为热尔曼素数。比如，5 是一个热尔曼素数，因为 $2\times5+1=11$ 仍是素数，而 7 就不是一个热尔曼素数，因为 $2\times7+1=15$ 不再是素数。

按照热尔曼的思想，勒让德推广了其结果，如果 p 是素数，使 $4p+1$、$8p+1$、$10p+1$、$14p+1$、$16p+1$ 之一也是素数，那么对于指数 p，费马大定理在第一种情形下成立。

这一推广后的结论实际上证明了对于所有素指数 $p<100$，第一种情形成立。沿着热尔曼开辟的这条途径，后来的数学家得到一些新的判断条件，从而大大提高了费马大定理第一种情形成立时的素指数。然而，这条途径是有局限的，即它只证明了在第一种情形中，相应指数的费马大定理成立。而对于更复杂的第二种情形下费马大定理的证明，还必须另辟蹊径。

大奖与暗礁

为了鼓励更多的人攻克费马大定理这座数学堡垒，也为了表示学术界对它的重视，1816 年法国巴黎科学院首次以费马大定理为题设置大奖，奖励由金质奖章和奖金组成。

听到消息后，高斯的朋友写信给数学之王：“亲爱的高斯，你应该为此忙碌一下。”

两周后，高斯答复说，自己对作为一个孤立命题的费马问题没有多少兴趣。这或许是高斯未能采摘下费马的葡萄而酸溜溜的自白吧。人们推测，高斯放弃费马问题，而把研究重点转向其他问题的一个原因或许是因为失败的苦涩：在他成功解决 $n=3$ 后，研究 n 为 5 以及 n 为 7 的情形时并没有成功。

虽然高斯放弃了这种诱惑，但更多其他数学家却投入到这种智慧的挑战中，并取得了新的进展。

1825 年，勒让德完整地证明了指数为 5 的情形。同年，年仅 20 岁的狄里克雷（1805—1859）在一篇论文中也试图证明指数为 5 的情形。但其证明不完全，这一点被勒让德指出，后来狄里克雷完善了自己的证明，并于 1828 年发表。

1832 年，狄里克雷解决了指数为 14 的情形。

1839 年，法国数学家拉梅（1795—1870）证明了指数为 7 的情形。不久，法国数学家勒贝格（1791—1875，不是后来创立测度论的那一位）给出了一个简化证明。

这些证明大都非常长，而且使用了非常困难而巧妙的处理，很难推广。要取得真正的突破，需要有全新的思路与方法。

1847 年 3 月 1 日，定期举行的巴黎科学院报告会上出现了具有历史意义的戏剧性一幕。

几年前完成指数为 7 的费马大定理证明的数学家拉梅走上讲台，宣布自己已通过一条新途径完全证明了费马大定理。其证明思路是，在复数域里有分解式

$$x^p+y^p=(x+y)(x+\zeta y)(x+\zeta^2 y)\cdots(x+\zeta^{p-1}y)=z^p$$

其中 p 是一个素数，$\zeta=\cos\dfrac{2\pi}{p}+\mathrm{i}\sin\dfrac{2\pi}{p}$ 是单位原根。

考虑被称为“分圆整数”的一类特殊形式的复数：$a_0+a_1\zeta+\cdots+a_{p-2}\zeta^{p-2}$，其中

a_0、$a_1 \cdots a_{p-2}$都是整数。容易知道，所有的分圆整数可构成一个数系。

然后，拉梅提出一个重要想法，在分圆整数系中，如果$(x+y)(x+\zeta y)(x+\zeta^2 y)\cdots(x+\zeta^{p-1}y)$这个式子中各个因子都互素，因为它们的乘积是$p$次幂的，那么每一个因子都是$p$次幂的。在此关键观点下，拉梅相信自己经过一系列步骤可以推得一般情形下的费马大定理成立。

拉梅做完报告后，在场的刘维尔马上站起来指出拉梅证明中存在的一个大问题。如我们所熟悉的，如果a、b互素，且ab是一个平方数，那么a、b本身都是平方数，进而，如果两个互素整数之积是一个n次方幂，那么每个数都应是n次方幂。这在整数范围内的确是可以证明的正确结论。然而，刘维尔指出，这一在整数中正确结论的前提是，任何整数都能被唯一分解成一些素数之积，即整数中唯一分解因子定理成立。但唯一分解因子定理在分圆整数系中是否还成立呢？这是非常值得怀疑的。

事实上，刘维尔正确指出了在费马大定理证明中存在一个暗礁，而它将使许多费马大定理的所谓数学证明触礁沉没。拉梅不是它的第一个也不是唯一的牺牲品。

刘维尔的发言刚结束，在座的另一位大数学家柯西就站起来表示支持拉梅，并声称他自己也有类似的证明思路，并已在1846年10月向科学院提交过一个报告了。

于是，证明分圆整数系唯一因子分解定理成为彻底解决费马大定理的关键。而最早证明这一点从而获得费马大定理完整证明的人将赢得数学中最权威的而且奖金丰厚的大奖。拉梅、柯西都意识到时间是至关重要的。此后许多天，两人在科学院内部发表了很多通报，并各自声明自己已在科学院存放了盖章密封的信封。这在当时是常有的做法，这能使研究者的思想被记录下来，而不泄露其工作的细节。如果后来关于谁最早有了某种想法出现争议，那么开启密封的信封就会判断出究竟谁拥有优先权。

就在人们纷纷猜测何人将获得最后的胜利之时，一封来自德国数学家库默尔的信

终结了这场无谓的竞争。这封信在 5 月 24 日的科学院会议上由刘维尔宣读，并震惊了所有参加会议的听众。

库默尔与他的大金蛋

库默尔（1810—1893），幼年丧父，家境困难，由母亲抚养长大。18 岁时，他进入哈雷大学学习神学。但在大学时他受到数学教师的影响，转攻数学，1855 年成为著名的柏林大学教授，直到退休。作为一名优秀教师，他对任何一种教学活动都很喜爱。在讲授任何一堂课前，他总是仔细备课，他的课讲得又清楚又有趣，就像说家常话，还带着泰然的微笑。库默尔生动幽默的教学风格，深深吸引了学生。此外，他还非常关心学生，并乐意帮助他们。很多故事都讲了他对学生的感情和关注。有一次，一个穷学生在博士考试前突然得了天花，年轻人不得不先考试，之后回了老家。库默尔没有那个不幸学生的消息，担心他在家里不能负担必需的医药费，就找到他的一个朋友，给他钱，让他去看看能帮什么忙。这一切使他赢得了学生极大的尊敬。在几十年教学生涯中，他培养了不少数学家。1882 年 2 月的一天，72 岁的库默尔以“记忆力衰退，思维能力减弱”为由提出退休。1883 年正式退休后，他完全远离了数学，在度过 10 年安静的生活后，于 1893 年 5 月 14 日在柏林去世。



库默尔

早在 1843 年，库默尔就对分圆整数进行了研究。研究中，他给出素整数、可除性以及类似东西的适当定义，并假定在分圆整数中唯一因子分解成立。由此出发，他给出费马大定理的一个证明。同年他把手稿寄给狄里克雷。狄里克雷看过后，指出他的假设没有一般性，即唯一因子分解仅对某些素数成立。1844 年，在认识到狄里克雷批评的正确性后，库默尔证明了分圆整数唯一因子分解定理不是普遍成立的。因此，当他看到拉梅的论文，并了解到拉梅与柯西的竞争时，他意识到这些数学同行遇到了同一块暗礁，并正在重蹈自己的覆辙。于是，他寄出了那封让当时“巴黎的饱学之士如雷轰顶、呆若木鸡”的信。在信中，库默尔首先提到“唯一因子分解定理在分圆整数系中是不成立的”，随信附上 3 年前发表的论文。此外，他又介绍了自己随后的研究成果。

在认识到唯一因子分解定理在分圆整数中不成立后，为了补救，库默尔定义了一种“理想数”。从理想数的观点看，因子分解的唯一性成立。随后，库默尔尝试用“理想数”证明费马大定理。为此，他提出了正则素数的概念。他证明了，对于所有的正则素数 p，费马大定理成立。对于哪些素数是正则素数，库默尔也给出了一个判断条件。根据这一条件，可以知道在 100 之内，只有 37、59、67 是非正则素数。这意味

着，根据库默尔的方法，在 100 之内的素数中，除去这三个例外，费马大定理都是成立的！这是历史上第一次真正对一整批素指数证明了费马大定理。想想在此之前，只是为了证明 n 为 3、5、7 就费去了多少精力，即可体会到库默尔的突破是何其巨大了。

1857 年，由于库默尔的这项开创性贡献，巴黎科学院把 1850 年第二次为征解费马大定理而设立的大奖颁发给了他，以奖励他“关于由单位根和整数组成的复数所做的美妙工作”。这对库默尔来说可谓是一个意外，因为他并没有把自己的研究成果交给巴黎科学院参赛，但在巴黎科学院看来，这是“做出的一项公正而有益的决定”。

正则素数的情况解决了，剩下的是非正则素数的情况。在这方面，还是库默尔本人迈出了第一步。1857 年，库默尔引入一个判据，这一判据可使得对某一类非正则素数而言费马大定理也成立。通过这一结果，费马大定理得以扩展到一些例外的非正则素数，特别是 100 之内的非正则素数 37、59、67 都属于这种扩展对象。因此，在库默尔的工作后，$p < 100$ 时费马大定理都被证明是成立的。但 60 年后，1917 年有数学家指出库默尔 1857 年的证明有错误。1926 年，美国数学家范迪威尔弥补了这些缺陷，并证明 $p < 157$ 时费马大定理成立。之后，在库默尔开辟的这条道路上，随着判据的不断改进，纪录不断被刷新。20 世纪 50 年代后，随着计算机的使用，纪录更是得以大幅度提高：

- ❑ 1929 年，$p < 211$；
- ❑ 1930 年，$p < 269$；
- ❑ 1931 年，$p < 307$；
- ❑ 1937 年，$p < 617$；
- ❑ 1954 年，$p < 2003$；
- ❑ 1954 年，$p < 2521$；

- 1955 年，$p < 4001$；
- 1967 年，$p < 25\,000$；
- 1975 年，$p < 30\,000$；
- 1975 年，$p < 58\,150$；
- 1976 年，$p < 100\,000$；
- 1977 年，$p < 125\,000$；
- 1987 年，$p < 150\,000$；
- 1992 年，$p < 1\,000\,000$；
- 1994 年，$p < 4\,000\,000$。

不过，这种推进上限的办法只是在创造一个又一个纪录。当人们证明了非正则素数有无限多时，库默尔开辟的途径被表明无法从根本上证明费马大定理。

然而，数学的发展表明，库默尔在解决费马大定理时引入的理想数理论与思想，其意义远超过了证明费马大定理本身。正如若干年后的 1900 年，希尔伯特在其著名的《数学问题》世纪演讲中所评价的："这样一个非常特殊、似乎不十分重要的问题会对科学产生怎样令人鼓舞的影响？受费马问题的启发，库默尔引进了理想数，并发现把一个分圆域的整数分解为素理想因子的唯一分解定理，这定理今天已被戴德金与克罗内克推广到任意代数数域，在近代数论中占据中心地位，其意义已远远超出数论的范围而深入到代数与函数论的领域。"

确实，库默尔理论在其产生后的百年（1870—1970）中，已由丑小鸭变成数学中最美丽的天鹅，它的辉煌使人目眩，而且在数学中处处都可以看到它的身影。费马大定理这个"会下金蛋的鹅"在库默尔手中下了一个真正的大金蛋。

第五节
第二次大突破

库默尔的伟大工作在费马大定理的证明中是一次真正的大突破。但这一次突破只能表现在量的积累上，并获得了一个又一个纪录，但它却无法在质上完成费马大定理的证明。在库默尔的工作后，证明费马大定理的前程仍是漫漫。人们还需要再寻其他途径。

10 万马克的奖金

到 20 世纪初，费马问题依然在数学家心目中占有特殊的地位。但在库默尔的工作之后，人们发现证明费马大定理的希望比以前更渺茫了。新一代数学家更多地转向各种不同的研究领域，不再理睬这看似不可能解决、只会让人徒劳无获的问题。迷人的费马大定理成为一个美丽却遥不可及的梦。然而，在 1908 年，一位德国人保罗·沃尔夫斯凯尔为这个问题注入了新的生命力。

沃尔夫斯凯尔是一位实业家，也是一位数学爱好者。没有人会预料到，

缺乏数学天赋的他会与费马大定理联系在一起。事情起因于一次失恋的经历。沃尔夫斯凯尔在追求一位漂亮女性时被拒绝了。为此感到极度失望的他放弃了生的念头，打算自杀。他定下了自杀的日子，并决定在那天的午夜钟声响起时开枪射击自己的头部。在剩下的日子里，他从容地处理各项事务。当他把身后的一切安排妥当时，他发现离自己定下的自杀时刻还差几个小时。百无聊赖之下，他开始翻阅数学书。结果，库默尔的经典论文吸引了他，他陶醉于库默尔的数学之中。当他意识到时间的流逝时，自杀时刻已经过去了。数学重新唤起了他生的欲望，他放弃了自杀。他去世后，人们从他的遗嘱中看到了他对挽救自己生命的费马猜想的报恩方式：他从自己的财产中拿出 10 万马克，交给哥廷根的皇家科学协会掌管，用以悬赏第一个证明费马大定理的人，期限 100 年（1908~2007 年）。

所有的数学杂志都刊登了这一新闻，消息迅速传遍欧洲。各地掀起了一股研究费马大定理的热潮。有当事人曾说在 10 万马克悬赏后，整个城市都沸腾起来，成千上万的教师、学生、商人、市民都被卷入其中。人们着了迷。论文像雪片似地飞到哥廷根大学。仅在 1908~1911 年间，就收到 1000 份以上所谓的证明。然而，毫无例外，所有应征者的方法和结论都是错误的。事实上，几乎所有的“解答”都是在非常初级的水平上（使用中学数学以及可能是数论中某些未经整理的论文中的概念）写成的，但尽管如此，理解起来却是非常复杂。

在 1909~1934 年间，哥廷根大学数学系的系主任朗道负责审查这些稿件。为了对付这件苦差事，他想出一个巧妙的办法。

他印制了几百张卡片，上面印着：

亲爱的＿＿＿＿：

谢谢您寄来的您关于费马大定理的证明的稿件。

第一个错误是在：

＿＿＿页＿＿＿行。

这使得证明无效。

E. M. 朗道教授

然后他把这些论文连同印好的卡片交给学生，要求学生填写空白处。

当时，德国有一个名为《数学和物理文献实录》的数学刊物，曾设立固定部门自愿对这方面论文进行鉴定。到 1911 年年初，该部门审查了 111 篇论文，全部发现错误。结果它最终也因无法承受审稿的沉重负担，而被迫宣布停止这方面的工作。

与热情的数学爱好者不同，尽管有宣传攻势和巨额奖金带来的额外刺激，专业数学家们却大多对此置之不理。因为他们非常清楚问题的难度，正如希尔伯特所认为的："为了做这件事，至少要花 3 年时间去做深入细致的研究，而到头来，还可能会失败。"征服费马大定理在当时的数学家看来，花费精力干这件事很可能是白白浪费时间。

费马大定理，这座横亘在数学家们面前的大山，仍然需要等待新的智者与新的思想来作它的征服者。这一等又是几十年，20 世纪 80 年代，代数几何领域出现的一个伟大定理意味着在费马大定理征程上出现了第二次重大突破。

一个伟大的定理

代数几何是解析几何的自然发展。在解析几何中，我们主要研究次数为一次、二次的代数曲线。到 19 世纪上半叶时，人们研究的对象扩展到三次或更高次的平面曲线，这就是代数几何的开端。

作为解析几何自然延续的代数几何，它与解析几何有一个主要不同点：解析几何用次数来对曲线和曲面分类，而代数几何则用一个称为“亏格”的不变量来对代数曲线进行分类。以亏格 g 为标准，所有代数曲线可以分为三大类：

- 直线、圆、圆锥曲线（$g=0$）；
- 椭圆曲线（$g=1$）；
- 其他曲线（$g\geqslant 2$）。

那么，费马大定理是如何与代数几何联系在一起的呢？很简单，方程 $x^n+y^n=z^n$（$n>2$）两端同除以 z^n，得 $\left(\frac{x}{z}\right)^n+\left(\frac{y}{z}\right)^n=1$，令 $x'=\frac{x}{z}$、$y'=\frac{y}{z}$，则有 $(x')^n+(y')^n=1$，这一方程定义的曲线称为费马曲线。

很显然，研究方程 $x^n+y^n=z^n$ 的整数解相当于研究 $(x')^n+(y')^n=1$ 的有理数解，或者说费马曲线上的有理点。因此，费马大定理等价于，在平面中，费马曲线在 $n>2$ 时没有坐标都是非零有理数点。由此数论问题与几何问题紧密联系在一起了。这种几何是考虑“有理点”问题的，也被称为“算术代数几何”或“算术几何”。

1922 年，英国数学家莫德尔（1888—1972）提出著名的猜想：亏格 $g\geqslant 2$ 的代数曲线上只有有限多个有理点。1929 年，西格尔证明了“有理点”是“整数点”的情形，

即亏格 $g \geqslant 2$ 的代数曲线上只有有限多个整数点。对于更困难的莫德尔猜想，数学界普遍认为它很难在短时期内得到证明。1972 年，在猜想提出整 50 年时，著名数学家魏伊对此评论说："……要是打赌的话，我愿赌它对而不愿赌它错，但这不过是一厢情愿，因为没有一点点支持或者反驳它的证据。"因此，当 1983 年这一猜想被证明的消息传出时，引起数学界极大的震动。

解决这一猜想的是刚得到博士学位 5 年的年轻德国数学家法尔廷斯（1954— ）。在花费了 18 个月的时间后，29 岁的法尔廷斯证明了这一猜想，得到了"一个伟大的定理"，同时他还解决了数学上另两个重要猜想：沙法列维奇猜想和塔特猜想。在证明中，法尔廷斯用了大量代数几何方面的最新成果。如有人评价的："从代数几何的整个发展来看，年轻的法尔廷斯有些像一个极富想象力的编花篮能手，将代数几何中许多漂亮的花朵……编织在两个骨架上，这两个骨架被数学家称为高度理论和 P 可除群理论。"其中高度理论是从费马的无穷递降法衍生而来的。



法尔廷斯

法尔廷斯的成就受到了高度评价，他证明莫德尔猜想的时间被称为"最近数学的一个伟大时刻"。1983 年 7 月 19 日《纽约时报》报道时，称法尔廷斯"解决了本世纪一个突出的数学问题""过去似乎是绝对解答不了的问题"。因这一世纪性成就，法尔廷斯于 1986 年获菲尔兹奖，年仅 32 岁。

法尔廷斯的伟大定理也是费马大定理证明的一大进展。在代数几何中，有结论：n 次光滑（即处处有切线）的曲线，其亏格 $g=\frac{(n-1)(n-2)}{2}$，n 代表多项式的次数。所以，1、2 次的曲线亏格为 0；光滑 3 次曲线的亏格为 1；光滑 4 次以上曲线的亏格 $g\geqslant 2$。考察由方程 $(x')^n+(y')^n=1$ 定义的费马曲线的亏格，可看到，当 $n\geqslant 4$ 时，费马曲线的亏格一定大于 1，因此利用法尔廷斯的伟大定理，得出费马曲线只有有限多个有理点。这等价于，如果方程 $x^n+y^n=z^n$（$n\geqslant 4$）有整数解，则至多有有限多个。这表明，费马大定理不成立的可能情形至多有有限个。因此，法尔廷斯的结果被公认为是费马大定理在库默尔之后最重要的进展，并被看作费马大定理的第二次重大突破。

法尔廷斯的突破，排除了费马问题无穷多解的可能性，下一步把这有限多个解降到只有平凡解，就可得到费马大定理，然而这看似简单的一步其实是很难跨越的。事实上，在法尔廷斯的结果后，没有多少数学家乐观地认为费马大定理会在短时间内得到彻底解决，他们认为离这座数学王国的珠穆朗玛峰之巅还有很长的距离。

此外，法尔廷斯的伟大定理还彰显出亏格为 1 的椭圆曲线的独特性。

椭圆曲线

椭圆曲线属于代数和数论中的对象，它已是当前数学研究的热点之一。椭圆曲线在数论中常用的形式是 $y^2=x^3+ax^2+bx+c$。这一名称容易望文生义地与椭圆看作一回事。事实上，椭圆曲线不是椭圆，椭圆是圆锥曲线，也就是二次曲线，而椭圆曲线是三次曲线。前者的亏格是 0，后者的亏格是 1。两者性质截然不同，它们之间只有极其

遥远的亲缘关系：椭圆曲线可用椭圆函数来参数化，椭圆函数来源于椭圆积分，而椭圆积分则来自求椭圆弧长，通过这一串链条，椭圆曲线才与椭圆挂上钩。

法尔廷斯的结论使人们对亏格 $g \geq 2$ 的代数曲线有了深入的认识：这类曲线只有有限多个有理点。对于亏格 $g = 0$ 的曲线，人们也早已认识清楚：这类曲线上或者没有有理点，或者有无穷多个有理点。于是，在法尔廷斯伟大定理得出后，算术代数几何中考虑的“有理点”问题只剩下椭圆曲线需要做深入探讨。

人们发现，看似简单的椭圆曲线事实上非常复杂。改变方程中 a、b、c 的值，会得到无穷多种方程，而每一种都有自己的特性。而其有理点数目问题也最为丰富。事实上，没有有理点、有有限多个有理点、有无穷多个有理点，这三种情形在椭圆曲线中都可能出现。人们也认识到，椭圆曲线上的有理点集有着极好的规律性：这些点构成一个阿贝尔群，并且这个群由有限多个元素生成。然而如何求得具体椭圆曲线的有理点或整数点呢?

300 多年前，费马曾向同时代数学家发出挑战：证明 26 是夹在一个平方数和一个立方数之间的唯一正整数。这相当于方程 $y^2 = x^3 - 2$ 只有一组正整数解。证明这一点是极其困难的，它曾难倒了费马的挑战对手。在费马之后，经过许多数学家的努力，不同椭圆曲线的解的个数仍是摆在数学家面前的极为艰难的课题。

一种简化的办法是，在“同余”的意义上寻找解。“同余”理论是高斯的伟大发明，通俗而言就是“时钟算术”。比如，我们熟知的钟表可看作 12 格时钟算术，其中 $9+5=2$，即 9 点再过 5 个小时是 2 点，还可以扩展做乘法，比如 $5 \times 7 = 11$，用同余式表示即 $9+5 \equiv 2 \pmod{12}$、$5 \times 7 \equiv 11 \pmod{12}$。经过这一简化后，人们发现对给定的时钟算术求出椭圆方程的所有可能的解相对而言比较容易完成。如在 5 格时钟算术中可以列出椭圆方程 $x^3 - x^2 = y^2 + y$ 的所有可能解，这些解是 (0, 0)、(0, 4)、(1, 0)、(1, 4)。注意，其中某些解在正规算术中是不正确的，但在 5 格时钟算术中却是对的。比如，

将 (1, 4) 这个解代入方程，可得到 0=20，这似乎是错的，但在 5 格时钟算术中这一等式是正确的。于是我们得到，对于这一给定的椭圆方程，它在 5 格时钟算术中解的个数是 4，数学家们说 $E_5=4$。如果换成 7 格时钟算术呢？数学家们发现同一方程的解的个数是 9，即 $E_7=9$。这样，数学家可以在各种不同的时钟算术中分别求出一个椭圆方程解的个数，并由此得到关于此椭圆方程的一个 E- 序列。如对于上面的方程，有序列 $E_1=1$、$E_2=4$、$E_3=4$、$E_4=8$、$E_5=4$、$E_6=16$、$E_7=9$、…对于数学家而言，这一序列浓缩了关于它所描述的椭圆方程的许多信息，相当于原椭圆方程的 DNA。

在对椭圆曲线的深入研究中，人们逐渐揭开了它的部分神秘面纱。与此同时，人们也深刻体会到这一曲线的极端重要性。在 20 世纪，人们发现它几乎无所不能，利用它可以得出许多重要成果，比如给出分解因子的有效方法等。对于我们要介绍的主题而言，它的重要性体现在一位椭圆曲线方面的专家以这种曲线为工具最终证明了费马大定理。

第六节
戏剧性的圆梦之旅

费马神秘撩人的页边评注为后世留下有史以来最著名的数学之谜，并由此引发了长达三个多世纪的充满惊险、悬疑、绝望和狂喜的数学探奇。这段历史先后涉及最多产的数学大师欧拉、数学之王高斯、柯西、库默尔、法尔廷斯……终于，怀尔斯（1953— ）出现了，他找到了谜底，把这出戏推向高潮并戛然而止，留下一段耐人回味的传奇。

童年梦想

许多年之后，面对台下的观众，怀尔斯将会想起坐在图书馆第一次遇到费马大定理的那一天。那时，怀尔斯还只是一个 10 岁的小男孩。在图书馆中，他正阅读 E. T. 贝尔的《大问题》一书。书中描述了费马猜想这头数学怪兽的存在。这是一个他完全可以理解的问题，然而他看到无数著名的数学家在这个难得出奇的问题面前束手无策，费马本人的神秘旁注以及 10 万马克悬赏的故事进一步激发了他的幻想。“从我孩提时代第一次遇到费马

大定理以后，它就一直是我最大的兴趣所在。”自那时起，怀尔斯就开始了证明费马大定理的探险。

还是孩子的怀尔斯像许多人一样相信费马证明的存在，单纯的他设想自己如果懂得的数学与费马一样多，那么他或许有机会找到费马那遗失的证明。于是，他从一个老师那里弄到一本数论方面的书，他掌握后就开始充满热情地尝试费马可能用过的方法。多次尝试都以失败告终后，他转向 18、19 世纪那些在费马的堡垒面前碰壁的数学家，他希望能从他们的错误中学到有用的东西。他理解了这些人的工作，但不得不在同样的障碍面前止步。

在 1974 年和 1977 年，怀尔斯分别获牛津大学学士学位与剑桥大学博士学位。“进入剑桥时，我真正把费马大定理搁在一边了。这不是因为我忘了它，而是我认识到我们所掌握的用来攻克它的全部技术已经反复使用了 130 年，而这些技术似乎没有触及问题根本。”因为担心耗费太多时间而一无所获，他把童年的梦想深埋心底，暂时放下了对费马大定理的思索。在导师科茨的建议下，他开始研究椭圆曲线理论——这个看似与证明费马大定理不相关的理论后来却成为他实现梦想的工具。

到 20 世纪 80 年代，怀尔斯在事业方面已取得了巨大成功。1982 年，他被任命为普林斯顿大学的教授。在数学界，这个安静而腼腆的人赢得了解题能手的声誉，并成为对代数数论做出巨大贡献的极少数优秀数学家之一。但在这些年中，他一直没有去触及曾萦绕于心的费马问题，深藏多年的童年梦想在等待着被重新唤醒的机会。不久后，改变他生命进程的时刻等到了。一座桥梁的出现将把他的童年梦想与他目前已有的专长联结在一起。事情要从几十年前说起。

桥　　梁

1954 年 1 月的一天，一个叫志村的日本年轻人搞研究时遇到一个极其复杂的计算，经过查阅文献他发现，有篇论文对他要进行的计算会有帮助。于是他急忙到系资料室去借阅有这篇论文的杂志。不巧的是，他需要的这一期杂志被人借走了，从借书卡上志村得知借者是自己不太熟悉的一位叫谷山的校友。志村写信给谷山，客气地询问对方什么时候可以归还杂志，并解释自己为什么急需这本杂志。谷山的回信令志村惊奇，原来谷山也正在进行同一计算，并且在同一处卡住了。于是两人戏剧性地相识了，从此开始了两人的合作。

在合作中，他们共同探讨了两人都非常着迷的模形式。模形式是数学中古怪和神奇的一部分，它的关键特点是具有非同寻常的对称性。研究中，人们发现每一种模形式都可由相同的一些基本要素构造出来。这些要素可以从 1 开始编号，一个特定的模形式可能包含 1 个 1 号要素、3 个 2 号要素、2 个 3 号要素等。这些刻画了模形式是如何构造的信息可以概括为模序列，或称 M- 序列，即要素及每一要素所需数量组成的表，比如 $M_1=1$，$M_2=3$，$M_3=2$，…。在数学家眼中，M- 序列是模形式的 DNA。在这一序列中，每个要素的数量起着关键的作用。

灵感在某一天光顾了谷山。他注意到有一个模形式的 M- 序列与一个椭圆方程的 E- 序列对应的数完全相同。这是意外的巧合吗？谷山又仔细研究了另几个不同的模形式。他发现，在每一种情形中，M- 序列似乎都完美地对应着某个椭圆方程的 E- 序列。1955 年，在进一步的思考后，谷山提出猜测：每一个模形式都有一个椭圆方程与之相对应，两者可以通过它们各自的序列发生联系。然而，模形式与椭圆曲线属于数学世界中完全不同的区域，没有任何明显的理由表明两者之间会存在如此紧密的联

系。为了使这一猜测更具说服力，谷山与志村合作，试图找到更多证据来支持存在于模世界和椭圆世界之间的这种联系。

1958 年 11 月 17 日，年仅 31 岁的谷山“因陷入对未来失去信心的心境之中”而自杀身亡。谷山去世后，志村继续收集支持这一猜测的更多证据。逐渐积累的证据使这一猜测逐渐被人们所承认，为表彰其提出者谷山和全力继续发展它的志村，它被称为谷山－志村猜想。这一使人震惊的神奇猜想断言：每个椭圆方程相伴着一个模形式，或者说有理数域上的所有椭圆曲线都是模的（即都可以模形式化）。如果这是正确的，那么在两个完全不同的数学领域（模形式与椭圆曲线）之间将搭起一座数学家们极其喜爱的数学之桥，从而把两者密切联系在一起。然而，没有数学家认为证明这座桥梁的存在是容易的。因此，几十年过去后，获得一个对这一漂亮猜测的证明对数学家们而言仍是遥不可及的事情。由于太超前于时代，谷山－志村猜想事实上在很长时间中被忽视了。直到它与费马问题之间搭起一座桥后，事情才发生了转折。

1985 年，在一次演讲中，数学家弗莱指出，如果费马猜想不成立，即费马方程存在一组非零整数解 a, b, c，使得 $a^n + b^n = c^n$ 成立，那么用这组解可以构造出如下的椭圆曲线 $y^2 = x(x+a^n)(x-b^n)$，这类椭圆曲线现在已被称为弗莱曲线。弗莱指出，这一曲线是如此复杂与古怪，因此它不可能被模形式化。

如果它真的不能被模形式化，那么谷山－志村猜想就是错误的，因为后者断言所有的椭圆曲线都可以被模形式化。反过来考虑，如果谷山－志村猜想是正确的，那么弗莱曲线就不可能存在。而它的存在是由假定费马方程有解导出的，这说明假定是错的，也就是说费马方程无解。当然，这就是费马猜想。简言之，若谷山－志村猜想成立，则费马大定理也成立。不过，在真正建立谷山－志村猜想与费马大定理的这一联系之前，还有一条小的逻辑缝隙：弗莱只是认为其给出的曲线不是模的，对此数学家需要更严格的证明。1986 年 6 月，美国数学家里贝特运用一个非常巧妙的想法填补了这一缺失的环节，他证明了弗莱曲线确实不能模形式化。由此，由谷山－志村猜想通

往费马大定理的通道被打开了，费马大定理与谷山－志村猜想联结在一起了。如果有人能证明每一个椭圆方程是模形式的，那么就可得出费马方程无解，于是费马大定理得证。于是，17世纪最重要的问题与20世纪最有意义的问题结合在一起了。一个在历史上和感情上极为重要的问题与一个可能引起现代数学革命的猜想联结在一起了。

即便如此，也没有人试图费神去证明谷山－志村猜想。绝大多数数学家相信，通过证明谷山－志村猜想来证明费马大定理，只是把一个几百年来未获解决的难题转化为一个看起来完全无法接近的猜想，这一条道路并非坦途。正如怀尔斯在剑桥的导师科茨所评价的："我自己对于这个存在于费马大定理与谷山－志村猜想之间的美妙链环能否实际产生有用的东西持悲观态度。"在他与其他众多数学家看来，"谷山－志村猜想不容易被证明。虽然问题很美妙，但真正证明它似乎是不可能的。我必须承认在我有生之年大概不可能看到它被证明"。试图冒险的只有怀尔斯。

谜底揭开

在一段时间里，费马猜想作为一个孤立的未解决难题与数学主流似乎是无关的。然而，谷山－志村猜想与费马猜想联系的建立，表明费马猜想已成为数学中一个不能不管的问题。怀尔斯是在一个朋友家饮茶时听到这个令他感到震撼的消息的，他后来回忆道："我记得那个时刻，那个改变我生命历程的时刻。"此后，怀尔斯放弃了所有与证明费马大定理无直接关系的工作，不再参加学术会议和报告会，开始了他秘密的数学冒险。他躲进他顶楼的书房，全身心投入到证明谷山－志村猜想中。

在顽强的探索中，他不断取得进展。对这种探索中的体验，他说："设想你进入大厦的第一间房间，完全是一片漆黑，你在家具之间跌跌撞撞，但是你逐渐搞清楚了每一件家具所在的位置。经过了六七个月，你终于摸到了电灯开关，打开了灯，突然一切都清楚了，你才确切知道你身在何处。于是你又找到隔壁的房间去，在黑暗中再摸索六七个月。每一次这样的突破，都是在黑暗中摸索几个月完成的。因此，这些突破都是摸索的结晶!"

当他被卡住时，他会沿着湖边散散步，并随时把笔纸带上，一旦有好主意会找个长椅坐下来打草稿。"问题整日整夜地在我的头脑里，"如他所说，"真是日复一日地干，几乎想不起来哪一天起床想的是别的事。"

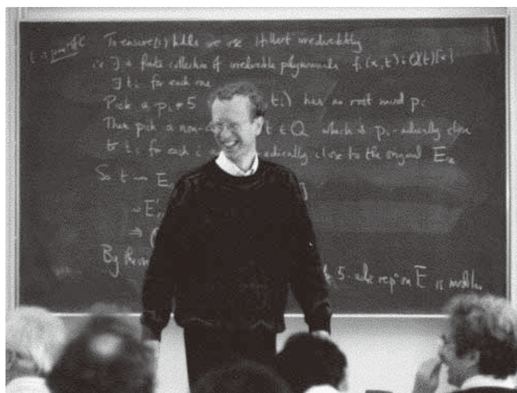
1993 年 5 月末的一个早上，怀尔斯的太太和孩子们都出去了。他坐在书桌前思考最后的步骤，不经意间读到了一篇论文，上面的一行字引起了他的注意。它提到了一个 19 世纪的数学结构，他霎时意识到这就是该用的。他不停地工作，忘记下楼吃饭。到下午三四点，在确信已经证明了费马大定理后他走下楼，告诉自己的太太："我解决了费马大定理。"

自 1986 年那特别的时刻起，已经过去了整整 7 年。在历经 7 年隐士般孤寂的岁月后，他相信自己已经实现了梦想。这年 6 月，他的导师科茨正好在剑桥牛顿研究所主持一个系列演讲。他决定在自己的家乡，在自己曾经就读过的地方，对外透露他的秘密研究。

1993 年 6 月，怀尔斯以"模形式、椭圆曲线和伽罗瓦表示"为题做了三次系列演讲。6 月 23 日，是第三次也是最后一次演讲。这时，消息已经传开。剑桥数学界的每一个人都来听这最后一次演讲。气氛充满了激情，兴奋的听众意识到他们正在跟随怀尔斯的足迹走向费马大定理的证明，他们正在参与一个历史性的事件。

报告从上午 8 点整开始，怀尔斯站在牛顿研究所的演讲厅里，不断在黑板上飞快

地写着，然后努力克制住自己的喜悦，凝视着台下的听众。10 点 30 分，在报告结束时，他最后一次转向黑板，写上了费马大定理的结论，然后转向听众，平静地宣布："我想我就在这里结束。"



相机闪烁拍下了这个历史性的时刻。会场上爆发出一阵持久的掌声。三百多年来，费马的挑战第一次被征服了。

虽然数学家们还不清楚证明的细节，但数学界普遍看好怀尔斯的证明。首先，他的证明途径被认为是可行的。为了证明费马大定理，他证明了部分谷山－志村猜想，即半稳定的椭圆曲线是模的，因为弗莱曲线恰是半稳定的椭圆曲线，因此这足以推出费马大定理。怀尔斯证明的整个结构看来也是十分可信和可靠的。此外，怀尔斯本人早已在此之前证明了自己的实力。而且作为一个非常仔细、从不草率宣布结果的人，他在同行中具有很好的口碑。已有的声望加上同行的看好，使人们确信：费马大定理这个貌似简单却曾使许多人求索而久攻不下的难题，终于土崩瓦解了。

消息很快轰动了全世界，怀尔斯成了全世界媒体的焦点。《纽约时报》在头版以《终于欢呼"我发现了!"，久远的数学之谜获解》为题报道费马大定理被证明的消息。其他各种大众传播媒介也都迅速报道了这一"世纪性的科学成就"。一夜之间，怀尔斯成为世界上最知名的数学家。

宣传的热潮一浪高过一浪。《人物》杂志将他与戴安娜王妃、克林顿总统等一起列为"本年度25位最具魅力者"。甚至一家国际企业要请这位腼腆的数学家为其新系列男装做广告！

与此同时，认真核对这个证明的工作也在进行。按照科学的程序，先要求数学家将稿件投到一个有声望的刊物，然后由刊物编辑部组织专家进行审核。每一个新的数学成果必须经过仔细的检查才能被接受是正确的。怀尔斯将手稿投到《数学发明》，编辑部组织了6位审稿人，并将分成6章的200页手稿交给6位审稿人，而每人负责其中的一章。怀尔斯在等待审稿人的意见中，焦急地度过了1993年的夏天。

8月23日，负责第3章的普林斯顿大学教授凯兹发现了一处缺陷。在如此长的证明中，出现错误并不意外。关键的是，这样的缺陷是否能弥补。第3章中出现的这一微妙的缺陷对怀尔斯而言究竟是瑕疵、裂缝、大深沟还是地狱？怀尔斯一开始很乐观地指望这是一处能很快解决的小麻烦，他希望能在数学界知道证明中有错误之前将这个错误改正过来。然而不久后他发现自己无法在短时间内完成修补。他重新回到与外界世界隔绝的状态，恢复了在顶楼研究的习惯，以填补这道裂缝。然而，随着时间的推移，问题似乎变得越来越棘手。1993年12月4日，怀尔斯在电子邮件中不得不承认"在检验过程中发现了许多问题，大部分已经解决，但是有一个特别的问题我还没有解决"。

历史似乎正在重复。人们回想起数学家宫冈洋一的失败。1988年3月，有实力的38岁日本数学家宫冈洋一沿另一方向曾给出费马大定理的一个证明。传闻尚未被证实，整个数学界已沸腾，世界许多媒体也都大肆宣扬，声称宫冈洋一证明了费马大定理，消息很快传遍了世界各个角落。然而，在随后仔细检查宫冈洋一的复杂证明时，人们发现了缺陷。更致命的是，并没有绕过困难的方法，补救工作宣布失败。从最初的声明算起两个月后，一致的意见是宫冈洋一的证明是失败的。于是，倒在费马问题上的有名数学家名单上又多了一位。不过，宫冈洋一的关于费马定理的证明尽管有漏

洞，但他的证明的整体规模宏大、旁征博引，具有非凡的知识广度及娴熟的代数几何技巧。因此，虽然他没有在费马猜想上取得成功，但是其证明中许多独特的部分具有其本身的存在价值，后来被别的数学家进一步发展，用于解决其他问题。

看起来，怀尔斯正在重蹈覆辙。他的证明也存在无法补全的缺口。他的有缺陷的证明对于了解他工作的那些数学家来说仍然是很有价值的。他新发展的一些数学技巧与技术，可以用来处理其他重要且困难的问题。只有一点，因为这一缺口的存在，怀尔斯的童年梦想仍然遥不可及。费马猜想仍然是横亘在怀尔斯面前的无法跨越的珠穆朗玛峰。

问题出在证明的关键一步上。为了解决这一关键环节，怀尔斯最初考虑使用伊娃沙娃理论，但经过两年的努力后，他走进了一条数学的死胡同。一年后，他发现了科利瓦金－弗莱切方法。他采用这种新技术并于 1993 年 5 月得到了自己的证明。但正是在使用后一方法的过程中，人们发出了一处隐藏很深的缺陷，一个隐藏在科利瓦金－弗莱切方法内部的漏洞。

为了越过障碍，在数学同行的建议下，怀尔斯邀请他以前的学生泰勒一起工作。然而任何修改科利瓦金－弗莱切方法的企图都失败了。因为一直拒绝公开手稿，怀尔斯不得不承受巨大的压力。1994 年 9 月，怀尔斯准备放弃自己看来无望的努力。泰勒建议再坚持一个月。就在截止日到来之前两周，9 月 19 日，一个星期一的早晨，“突然间，不可思议地，我发现了它……它美得难以形容，简单而优雅。我对着它发呆了 20 多分钟。然后我到系里转了一圈，又回到桌子旁边看看它是否还在那里——它确实还是那里。”克服最后障碍的途径找到了，怀尔斯发现了解开症结的戏剧性方法。他意识到，虽然科利瓦金－弗莱切方法不能完全行得通，但是只需要它就可以使原先尝试过但又放弃的伊娃沙娃理论奏效！他认识到科利瓦金－弗莱切方法中有足够的东西使他原先对这个问题的处理方法取得成功。所以，正确答案似乎在科利瓦金－弗莱切的废墟之中。单靠伊娃沙娃理论不足以解决问题，单靠科利瓦金－弗莱切方法也不够。

它们结合在一起却可以完美地互相补足。这是怀尔斯永远不会忘记的充满灵感的瞬间，他激动得泪水夺眶而出。在回忆这一时刻时，他说："我无法控制自己，我太兴奋了。这是我工作经历中最重要的时刻……" 在经历了他数学生涯中充满了痛苦、羞辱和沮丧的 14 个月后，一个高明的见解使他的苦难走到了尽头。

1994 年 10 月 14 日，怀尔斯将修改后的 108 页论文《模曲线和费马大定理》送交当代最权威的数学杂志——普林斯顿的《数学年刊》，而一周之前，他和他的学生泰勒的合作论文《某些海克代数的环论性质》已经寄去审查。第一篇论文宣布了费马大定理的一个证明，它的关键一步有赖于第二篇的合作论文。新的处理方法比原来的更简单和快捷，两篇论文总共有 130 页。在经过严格而彻底的核查后，1995 年 5 月，《数学年刊》用整期篇幅发表了他的论文。证明现在是可靠的了。怀尔斯再一次出现在《纽约时报》头版，文章标题是《数学家称经典之谜已解决》。

尘埃落定。一座坚硬无比的数学堡垒被攻克了。扮演"费马大定理终结者"角色的怀尔斯 8 年潜心努力终获完满回报。他也终于在成年时期圆了童年时代的梦想。他的成果被认为是"20 世纪最重大的数学成就"。



怀尔斯的证明对数学家而言不仅仅是登上了费马大定理这座高峰。事实上，一个人可能找到一个证明，但这个证明可能没有什么深远的意义。不同的是，怀尔斯的证明确确实实地孵出了金蛋，其重要性超过了费马猜想本身。

首先，怀尔斯在证明中创造了全新的数学技术，并将它们和传统技术以人们从

未考虑过的方式结合起来。通过这样的做法，他开辟了处理众多其他问题的新思路，用他的新技巧可以证明更多的东西，从而对未来产生不可估量的影响。举一个例子，2001 年几位数学家按照他的研究策略和方法证明了所有的椭圆曲线都是模的，从而彻底解决了谷山－志村猜想。

此外，怀尔斯的这个证明是现代数学的完美综合。它综合利用了现代数学许多分支（椭圆曲线论、模形式理论、伽罗瓦表示理论等）的成就，特别是代数几何领域中关于椭圆曲线的深刻结果。通过证明费马大定理，怀尔斯将一些最重要的数论发现凝聚在一起，把相隔很远的数学领域结合在一起，并扩展了数学王国的整个边界。此定理的攻克也再次显示了数学大厦的统一性。

另外，他的证明也完美体现了数学魅力的一面：一个小学生能听懂的问题竟蕴含如此深奥的原理，并要用到最现代化的数学工具才最终解决。

正因此，怀尔斯的证明为他赢得了最慷慨的褒扬，其中最具代表性的是他的导师科茨的评价："用数学的术语来说，这个最终的证明可与分裂原子或发现 DNA 的结构相比。证明费马大定理是人类智力活动的一曲凯歌，同时，不能忽视的事实是它一下子使数论发生了革命性的变化。对于我来说，怀尔斯的成果的美和魅力在于它是走向代数数论的巨大一步。"

此后几年，荣誉纷至沓来。

1996 年，由于在数论及相关领域的杰出贡献，特别是证明了费马猜想，怀尔斯荣获沃尔夫奖。这一著名的数学大奖是一种终身成就奖，获奖者一般要七八十岁才能获得。怀尔斯获奖时年仅 43 岁，是迄今为止获得此奖的数学家中最年轻的一位。

1996 年，怀尔斯获奥斯特洛夫斯基奖，这是欧洲最大的一项数学奖。

1996 年，怀尔斯获美国国家科学院数学奖。

1996 年，怀尔斯被选为美国国家科学院外籍院士。

1997 年 6 月，怀尔斯获得沃尔夫斯凯尔奖金。10 万马克的奖金因为一战后德国恶性通货膨胀大为贬值。到 1997 年颁奖时，奖金加上利息大约还值 5 万美元。

1998 年，国际数学家大会召开。会议的一项重要内容是宣布菲尔兹奖得主名单。按照惯例，这一奖项只授予不超过 40 岁的数学家。当时怀尔斯已经 45 岁了，但为了表彰他的光辉成就，大会为他颁发了菲尔兹特别贡献奖，这是迄今为止唯一荣获此奖的 40 岁以上的数学家。当他上台领奖时，人们给予他的掌声比其他 4 位菲尔兹奖得主更为热烈、持久。

2005 年，怀尔斯获邵逸夫奖，以及 100 万美元的奖金。

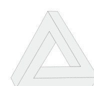
结　　语

一场旷日持久的猎逐就此结束，一出宏大的数学戏剧落下了帷幕。随着费马大定理的最终获证，这个困惑了世间智者 358 年的谜被揭开了。然而，费马是否真的得到了一个美妙的证明呢？这个谜中之谜没有随着费马大定理的获证而尘埃落定。现在可以肯定的一点是，费马绝对不可能使用怀尔斯的方法来解决这一难题。因为怀尔斯所使用的数学工具都是费马逝世后多年才逐渐发展起来的。不过，这一谜中之谜的谜底似乎不再重要。因为无论谜底如何，作为伟大的创造者，费马早已跻身于世界一流数学家的行列并赢得了“业余数学家之王”的美誉。而对数学而言，真正重要的是由此引出的费马猜想“扮演了类似珠穆朗玛峰对登山者（在成功之前）所起的作用，它是一

个挑战，试图登上顶峰的愿望刺激了新的技巧和技术的发展与完善”。

1908 年，沃尔夫斯凯尔奖设立后，有学生曾问当时最有可能解决这一问题的大数学家希尔伯特，为什么不攻下费马大定理而取得 10 万马克的奖金？希尔伯特一语双关地回答：“为什么要去杀死一只会下金蛋的鹅呢？”这句话可以有两种解释。一种解释认为，在沃尔夫斯凯尔的遗嘱中规定 10 万马克奖金发出之前，其利息由一个委员会决定如何使用。委员会主席希尔伯特决定用这笔不少的钱（年息 5000 马克）进行学术研究与交流，特别是邀请著名数学家与物理学家到哥廷根访问。这样，10 万马克的利息就可以不断为发展数学做出贡献，要是问题解决了，也就一无所有了。不过，现在人们更倾向于另一种解释，在解决问题的过程中，不断给数学加进崭新的内容，这些新内容对数学发展的意义甚至超过原来一个特殊问题的解决。

确实如此。一代又一代的数学家在证明费马猜想的过程中，在数学王国里挖掘出无数的宝藏。这只“会下金蛋的鹅”，给整个数学带来了巨大财富，促进了代数数论和算术代数几何学的建立，还发展了一系列先进数学技术，形成了现代数论无尽的前沿。



第六章

素数问题

第一节
素数

“素数是那些除了其自身和 1 之外无法被任何整数整除的、令人恼火的、不守规矩的整数。”在很早的时候，素数就成为数学的一项重要研究内容。现在，数学家尤其是数论专家仍然对素数表现出特别的青睐。是什么原因使得素数显得如此重要？素数又是如何不守规矩的呢？本节就来介绍一下。

素数的地位

一个大于 1 的自然数，如果只能被 1 和它本身整除，则称为素数，或者叫质数。一个自然数除了能被 1 和它本身整除外，还能够被其他自然数整除，则称为复合数，或合数。在这种分类方式下，我们得出自然数分三类：特殊的 1、素数、合数。

素数在自然数中是特殊的一类，又是极其重要的一类，它在数学中有着非常独特的位置。事实上，在具有固有特性的各类数中，它被研究得最

多。为什么素数如此重要？这建立在下面这个重要结论上：任何一个大于 1 的自然数都可以分解成几个素数的连乘积的形式，用数学式子表示，就是 $a = p_1 p_2 p_3 \cdots p_r$，$r \geqslant 1$。如果不计顺序，这种分解是唯一的。因为式子中的各个素因子可能相等，所以上式也可以写成标准分解式的形式：

$$a = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k},\ a_1,\ a_2,\ \cdots,\ a_k > 0$$

这一结论（称为唯一分解定理）在数论中非常重要。自然数中的许多重要性质与定理都以它为基础，因此它也被称为“初等数论基本定理”。因为数论又叫高等算术，或者简称算术，于是这一结论也习惯称为“算术基本定理”。它在数论中的作用由其名称可见一斑。

算术基本定理的存在，把素数推到了一个极为重要的位置上。这一事实表明，素数是所有自然数的“基础材料”，所有的自然数（除 1 外）都是由它们构成的。因而，我们可以非常贴切地把素数看作自然数的基石。事实上，只要掌握了任一个数的素因子分解，数学家就获得了这个数的众多信息。从这一角度上看，素数在数学中的地位非常类似元素之于化学或基本粒子之于物理。作为建筑自然数大厦的砖石，素数在数学，尤其是在数论中起着非常重要的作用。

素数的个数

在素数研究中，我们面对的第一个首要问题是，素数的个数是有限的还是无限

多？早在 2000 多年前，欧几里得就在《几何原本》第九卷中给出了答案：任意给定几个素数，总有与它们不同的素数存在。这一结论意味着，素数有无限多。

我们用反证法证明一下。假设素数只有有限个。现在把它们由小到大排列起来，并设其为 p_1，p_2，p_3，p_4，⋯，p_k，其中 p_k 为最大的素数。现在令 $N = p_1p_2p_3p_4\cdots p_k + 1$ 。下面我们来考察 N。

一种情况是 N 是素数，那么显然 N 大于 p_k，于是假设是错误的。

另一种情况是 N 是合数，此时 N 有素因子。由 N 的构造容易发现，p_1，p_2，p_3，p_4，⋯，p_k 都不可能整除 N，因为它们除 N 时都会余 1。所以，p_1，p_2，p_3，p_4，⋯，p_k 这些素数都不可能是 N 的素因子。因此，N 有不同于上述有限素数的新的素因子。这与假设相矛盾，所以假设不成立。无论何种情况，假设都是错误的。因此，素数有无限多。

欧几里得在《几何原本》中给出的证明是用几何语言表述的，但其实质与上面的证法相同。欧几里得的证明简洁、优美，又极为深刻，因此成为数学论证的经典之作，被誉为数学史上永铭的典范。20 世纪英国数学家 G. H. 哈代在其精彩著作《一个数学家的自白》中称欧几里得的证明“……自发现之日至今，永葆其生机与效力，两千年岁月没有使它产生一丝陈旧感”。

欧几里得的这一证明，除了成为数学美的例证为人所铭记外，他在此命题证明中所使用的证明方法更是产生了深远的影响，至今还在数学王国中显示出它的强大威力。

一方面，欧几里得对这一定理的证明采用了反证法。哈代曾高度评价这一方法，他说：“欧几里得所钟爱的反证法是数学家拥有的最好武器之一，它比任何一种着棋的弃子法要高明。棋手可能牺牲一卒、一子，然而，数学家却牺牲掉整盘的棋。”这种巧妙的证明方法在代数、几何、三角、数论等几乎所有的数学分支中都有广泛应用。它

对于解决否定性问题、不存在性问题、无限问题以及等量或不等量问题，常常是很有效的。

欧几里得的证明还妙在，它仅向人们宣称有某一珍宝存在，但没有泄露它在什么地方。这实际上是开了纯粹存在性证明方式之先河。

除了方法上的优美之外，欧几里得这一定理所表述的内涵对素数的研究也有着开创性意义。正是这一结论开创了素数个数问题的研究，引出了素数分布等一系列重大的研究课题。

素数寻踪

自欧几里得时代将素数看作数学中的“原子”，并引发了数学家们的研究之后，素数的寻找就成了重要的探讨课题。是否存在寻找素数的理想方法？自欧几里得之后，数学家们开始上穷碧落下黄泉，探寻素数的秘密。

公元前 250 年左右，古希腊数学家埃拉托塞尼（在第二章中我们已经介绍过他）创造了著名的古典筛法。其具体做法是，从 2 开始，将自然数从小到大的顺序写出来。2 是第一个素数，留下，将 2 后面所有 2 的倍数划去；2 后面第一个未被划去的是 3，留下，然后将 3 后面所有 3 的倍数划去。同理，保留 5，划去 5 的所有倍数……如此不断地重复进行下去。这样合数就好像从“筛子眼”里一批一批地漏掉了，剩下的未被划去的数就是自然数中的素数。按照埃拉托塞尼筛法，人们可构制出素数表。1657 年，有人构制了 10 000 内的素数表。1746 年，这一纪录被改写为 100 999。1772 年，

纪录被进一步提高到 400 000。1914 年，美国数学家 D. N. 雷默编订出从 2 到 10 006 721 的素数表，此表长期作为研究素数的工具，十分著名。1950 年后，计算机的问世使素数表的上限有了更大提高。1959 年，美国的贝克等人使素数表突破 1 亿大关。1976 年，贝斯和赫德森计算出直到 1.2×10^{12} 的素数表。随着计算机的迅速发展，造更大的素数表已不再是什么难事。

从原则上说，埃拉托塞尼筛法解决了有限的不太大范围内素数的寻找问题。事实上，迄今为止，这种筛法仍然是找出某一不太大范围内所有素数的最为迅速的方式。计算机的出现与发展，使其变得更为容易。

但这种方法的局限在于，它只能找出不太大范围内的素数，而且它在操作中所涉及的仅是有限问题。因此，在素数寻找中，还极有必要寻觅其他途径。其中一种努力方向是找到能产生素数的某些特殊方法，或者说寻找素数的表达公式。

在数论领域留下不可磨灭足迹的费马沿此方向迈出了第一步。正如在上一章中我们已经介绍过的，费马在 1640 年对 $2^{2^n}+1$ 这类数做了探讨。他发现当 n 取 0、1、2、3、4 时，对应值分别为 3、5、17、257、65 537，这几个数都是素数。于是，费马提出形如 $2^{2^n}+1$ 的数一定为素数。对费马所研究的 $2^{2^n}+1$ 这种形式的数，后来人们称为费马数，并用 F_n 表示。费马当时的猜测相当于所有费马数都一定是素数。然而，1732 年，欧拉发现 F_5 是一个合数，因此费马的猜想是错的。随后的研究表明，非常可能的情况是，在所有的费马数中，除了前 5 个是素数外，其他的都是合数。

另一类著名的素数称为梅森素数。1644 年，梅森出版了《物理学与数学的深思》一书。他在书的序言中写道："总结前人的工作和我个人的研究，可以得到结论：在 $n\leqslant257$ 的数中，除了当 n=2、3、5、7、13、17、19 时，2^n-1 是素数外，并猜想 n=31、67、127、257 时 2^n-1 也是素数。而 $n<257$ 的其他数值，2^n-1 都是合数。"梅森的猜想有许多错误，但人们为了纪念他并对他的努力和成就表示敬意，还是把形如 2^p-1（p 为素数）的数称为梅森数，记作 M_p。如果 M_p 是素数，则其称为梅森素数。

历史上许多人曾为寻找更大的梅森素数而努力。计算机出现后，有关最大梅森素数（往往也是最大的素数）的纪录不断被刷新。然而，p 是什么素数时，梅森数恰好是梅森素数呢？对此，却没有简单的法则可以遵循。

在寻找素数公式方面，著名数学家欧拉做出了另一种尝试，他开辟了用二次多项式表示素数的新途径，并且给出一个非常美妙的多项式 $f(n)=n^2+n+41$ 。这个式子具有如下性质：如果让 n 取 0, 1, 2, ⋯, 39 中的任一自然数，代入上面的式子，可以发现得到的数值总是素数。如 $f(0)=41$, $f(1)=43$, ⋯, $f(39)=1601$ 都是素数，这样共得到了 40 个不同的素数。但当 n 取 40 时，公式开始失效，得到的数不再是素数。也就是说，欧拉多项式无法总是给出素数。

于是我们看到，经过不懈努力后，数学家找到了一些能表示素数的公式，然而这些公式没有一个能总是产生素数。真正有用的、理想的产生素数的公式仍遥不可及，或许根本就不存在这样的公式。

素数的分布

在素数问题中，另一个研究方向是探讨素数的分布，这可以说是素数性质研究中也是数论中最重要和最有吸引力的中心问题之一。在这方面，一种探讨途径是研究具有某种特征的素数分布情况。沿这条途径，一个少见的一般性结论由德国数学家狄里克雷获得。

我们已经知道素数有无穷多。那么对于具有某种特征的素数来说，情况又如何

呢？在这类问题中，较简单的是形如 $4n-1$、$4n+1$、$3n+2$、$6n-1$ 等的素数个数问题。不难证明上述这些类型的素数都有无限个。于是，我们可以再推进一步，考虑一个一般性的问题：在等差数列 $km+l$，$m=0, 1, 2, \cdots$中是否有无穷多个素数？显然，当 $(k, l)\neq 1$ 时，数列中的素数不可能无穷多。$(k, l)=1$ 时，情况又如何呢？法国数学家勒让德曾提出猜想：在首项和公差互素的算术级数中，存在着无限多个素数。这个猜想勒让德本人没能解决。1837 年 7 月 27 日，狄里克雷向柏林科学院提交了一篇论文，在论文中他引进狄里克雷 L 函数并圆满证明了勒让德的猜想。这一结果被称为等差数列的素数定理或狄里克雷定理。重要的是，狄里克雷的证明还开创了应用解析方法证明数论问题的新途径，促进了解析数论的诞生。

另一种途径则是探讨整个素数在自然数中的分布情况。在探讨整个素数的分布规律中，我们首先想了解的是，在自然数列中，能不能确定在多大范围内，一定会有一个素数？如果考虑固定区间内的素数分布情况，得出的结论或许多少有点出人意料。事实上，不管这个固定区间有多大，我们都可以让这个区间内一个素数都不存在。换一种说法，两个素数之间的距离可以要多远有多远。证明这一点也不困难。

设 N 是任意自然数。我们把数 $(N+1)!=1\times 2\times 3\times 4\times\cdots\times(N+1)$ 记作 H。

显然，$H+2, H+3, \cdots, H+N, H+(N+1)$，这 N 个相邻的数都是合数而非素数。于是，只要 p 是 $H+2$ 前面最近的那个素数，而 q 是 $H+(N+1)$ 后面最近的那个素数，那么 p 与 q 之间超过 N 个数都不是素数。

因此，不管 N 有多大，也就是说给出的固定区间有多大，我们总可以找到两个素数，使其在如此长的固定区间内不存在一个素数。素数之间的间隙竟然要多大有多大！

然而，在另一方面，许许多多的素数又会紧挨在一起，成对地出现。这就是孪生

素数对问题。所谓孪生素数对，即 3 与 5、5 与 7 这样的相邻素数对，其定义是，若 p 与 $p+2$ 两数都是素数，那么我们把这两个数称为双生素数对或孪生素数对。人们已经发现许许多多的孪生素数对。一个令数学家感兴趣的理论问题是，孪生素数对是否有无穷多？这就是著名的孪生素数对猜想。这一问题是素数理论中的重要问题之一，迄今尚未解决。

第二节
素数定理

通过上面的介绍，我们看到素数行踪不定，令人难以捉摸。没有任何神奇的公式可以确切地告诉你哪些数是素数，哪些数不是素数。而在素数分布方面，一方面，素数之间可以相隔任意远，其间隙可以要多大有多大；另一方面，间隔小到 2 的孪生素数对可能有无限多！素数的分布就局部来说真是变幻莫测，不可捉摸。那么，素数的分布真的没有一点规律可循吗？一个先被猜测后被证明的著名结论表明素数的出现还是有一定的数学模式的。

素数定理

如果我们放弃细节处的变化，考虑整体变化，是否可以发现一些素数分布的总体规律呢？是否存在一个公式，它能给出小于 x 的素数个数？或者说能否找到一个小于指定上限的素数个数的近似公式？

对此的研究是通过当时已有的素数表进行的。人们可以借助素数表数

出在某个范围内的素数个数，然后从经验中概括出猜想，由此探寻在某个范围内素数个数可能遵循的模式。比如，我们可以数出在 1 到 10 之间有 4 个素数，在 1 到 100 之间有 25 个素数，在 1 到 1000 之间有 168 个素数……数论中经常用 $\pi(x)$ 表示不大于 x 的素数的个数。于是有 $\pi(10)=4$，$\pi(100)=25$，$\pi(1000)=168$……在有了更大的素数表后，我们还可以给出更大范围内的素数个数，如下所示：

x	$\pi(x)$
10	4
100	25
1000	168
10 000	1229
100 000	9592
1 000 000	78 498
10 000 000	664 579
100 000 000	5 761 455

通过数素数表中素数的个数，能发现什么呢？比较明显的一点是，素数间的“跨度”会越来越大，或者说素数越来越稀少。然而具体到某个范围，素数的分布仍然显得变幻莫测。在某些范围内素数可能很多，但在另一个范围内就可能完全没有素数。而在某个固定的范围内，素数的数量既非连续不断地递增，亦非递减，而是杂乱无章地变动。比如，同是间隔 100 个连续的自然数，在 5800~5900 有 16 个素数，在 5900~6000 仅有 7 个素数，在 6300~6400 多达 15 个素数，而 6400~6500 仅有 8 个素数……有人因此把素数描述为在自然数之间随机生长的野草。确实在许多世纪中，数学家们一直未能成功地说明素数的构成模式。伟大的欧拉也认为素数的分布是“人类心智永远不能看透的谜”。看来，要想看透素数分布的谜，除了勤奋地用平凡的计算去识别素数和计算素数的个数外，还需要深邃的洞察力。重大突破终于在 18 世纪末出现。

1798 年，法国数学家勒让德首先公布了自己的发现。在对素数分布观察后，他用归纳法得到一个素数分布定律：$\pi(x)$ 大致等于 $\frac{x}{\ln x-1.08366}$。这一公式也记录在其《数论》(1808 年) 一书中。这一公式中的 1.083 66 看起来有些奇怪。事实上，这个数值是勒让德在检查了 400 000 以内的素数表之后得到的，选择此数只不过是为了给出尽可能好的结果。因此这是一个名副其实的经验公式。阿贝尔在 1823 年的一封信中称这个结论是“在整个数学中最值得注意的”。

然而，勒让德的发现刚公布，数学之王高斯的声明随之而至。高斯说自己早已在几年前就已发现了素数个数函数与对数函数之间存在的这种奇妙联系。由此导致了两人之间一场激烈的论战。如我们已经知道的，高斯习惯于把自己的发现秘而不宣，或只是在与朋友的私人通信中做一些透露，而不幸的是高斯与勒让德两人的研究工作非常相似，结果两人的争执发生了不止一次。直到高斯去世后，他的通信被公开，人们才确信在这些争论中高斯确实是领先的一方，如在素数分布规律方面，高斯的发现比勒让德早了 6 年。

1791 年，14 岁的高斯得到一份礼物——一本有关对数的书，其中书后附有一张素数表。1792 年，15 岁的高斯在对素数的研究中，敏锐地注意到“对数”与“素数”这两种看上去毫不相干的事物之间似乎存在着某种联系。他的发现是，素数分布的密度差不多与其自然对数成反比，即素数的分布密度可近似地用 $\frac{1}{\ln x}$ 来描述，由此小于 x 的素数个数 $\pi(x)$ 可粗略估计为 $\frac{x}{\ln x}$。

作为在不大于整数 x 的任意大整数范围内素数个数的估算公式，高斯最初给出的 $\frac{x}{\ln x}$ 是粗糙的。当 x 变得越来越大时，这个估算与真实值的绝对偏差会越来越大。相比而言，人们认为勒让德给出的 $\frac{x}{\ln x-1.08366}$ 效果要好一些。这并不奇怪，因为勒让德所做的正是通过修改公式来模拟已知数据。但随着更大范围素数表的建立，人们发现勒让德的公式对于超过 10 000 000 之后的素数个数也根本谈不上精确。而且公式中看上去相当粗糙的修正项 1.083 66，使得数学家相信肯定有更好、更自然的东西隐藏

在背后。确实，不久后高斯就得到了一个更好的改进结果，他给出一个对数积分函数 $\mathrm{Li}(x)=\int_2^x \frac{\mathrm{d}t}{\ln t}$ 作为 $\pi(x)$ 的估计公式。这一结果不仅形式上更加优美，而且结果也十分精确。

下表对照给出了几种公式的估算效果。

x	$\pi(x)$	$\frac{x}{\ln x}$	$\frac{x}{\ln x-1.08366}$	$\mathrm{Li}(x)$
1000	168	145	172	178
10 000	1229	1086	1231	1246
100 000	9592	8686	9588	9630
1 000 000	78 498	73 382	78 543	78 628
10 000 000	664 579	620 421	665 140	664 918
10 000 000	5 761 455	5 428 681	5 768 004	5 762 209

现代的素数表可以更好地证实高斯的 $\mathrm{Li}(x)$ 可以更好地估计 $\pi(x)$。比如，对 10^{16} 以内的素数个数，高斯的估计值与真实值只相差了十亿分之一，而勒让德的估计值则相差了千分之一。

在得到这一相当准确的结果后，通过观察，高斯发现了一个有趣的现象：$\mathrm{Li}(x)$ 比 $\pi(x)$ 的值要偏大些，它似乎总是稍稍高于真正的素数分布情形。对不大于 100 万、10 亿或 10 000 亿的素数进行测试，都显示出高斯的公式有点慷慨。这强烈地诱惑包括高斯在内的数学家们相信这种情形对直到无穷的一切数都是对的，从而产生了高估素数猜想，即在任何情况下总会有 $\pi(x)<\mathrm{Li}(x)$。一段时间内，人们相信这一猜想是正确的。然而，1914 年，英国数学家李特伍德证明了，在充分大的数字范围内时高斯的公式将会低估素数的个数，而且这种颠倒会发生无穷多次。观察欺骗了我们。1955 年，斯古斯显示这种低估在到达数字 $e^{e^{e^{79}}}$ 之前某处出现。这个数约相当于 $10^{10^{10^{34}}}$。这个难以想象的大数，被叫作斯古斯数，它可谓是“数学中迄今为止为确定目的服务过的最大数”。

$\mathrm{Li}(x)$ 作为 $\pi(x)$ 的一个估计公式已经非常好了，那么它是否还可以进一步改进呢？另外，一个重要问题是，$\pi(x)$ 与 $\mathrm{Li}(x)$ 之间的误差究竟有多大呢？在下一节中，我们将看到数学家特别是极具独创性的黎曼对此做出的回答。

回到高斯与勒让德的猜测，他们的工作实质上关注的都是素数分布的一种总体趋势，即考察 x 越来越大时 $\pi(x)$ 的渐近趋势。而在渐近意义上，$\frac{x}{\ln x}$、$\frac{x}{\ln x-1.08366}$、$\int_2^x \frac{\mathrm{d}t}{\ln t}$ 是完全等价的。因此，两人的猜测都可用极限语言表述为 $\lim\limits_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}}=1$，即当 x 趋近于无穷大时，$\frac{\pi(x)}{\frac{x}{\ln x}}$ 趋近于 1，或者说，x 趋近于无穷大时，$\pi(x)\sim\frac{x}{\ln x}$。

然而，在做出证明之前，人们还不能确定这样的素数分布法则是否正确，数学家需要找到这一猜想的证明。

在证明方面，首先迈出重要一步的是俄国数学家切比雪夫。他在 1852 年左右，取得了一项重要成果。他证明了存在两个正数 A_1 和 A_2，使得不等式

$$A_1\frac{x}{\ln x}<\pi(x)<A_2\frac{x}{\ln x},\ x\geqslant 2$$

成立。

这就是著名的切比雪夫不等式。事实上，切比雪夫也正是从这一结果出发，证明了素数理论中第一个漂亮的结果：贝特兰猜想。这一由法国数学家贝特兰于 1845 年提出的猜想是，对于大于 1 的自然数 N，在 N~$2N$ 之间，至少存在一个素数。

切比雪夫迈出了第一步，但这一步离证明高斯与勒让德关于素数分布的猜想还有不短的距离。当时一位数学家赛尔凡斯特曾对这一猜想的前景做出展望："但是要确定这种可能性的存在，我们或许要等待世界上出现这样一个人，他的智慧与洞察力像切比雪夫一样，证明自己是超人一等的。"

就在赛尔凡斯特说这些话时法国数学家阿达玛出生了，他在 1896 年证明了当 x 趋近于无穷大时，$\pi(x) \sim \dfrac{x}{\ln x}$。比利时数学家瓦莱·普桑（1866—1962）在同一年也独立证明了这一结果。

法国著名数学家阿达玛（1865—1963），在多个数学领域都做出过重要贡献。他还完成了流传很广的《数学领域中的发明心理学》。



阿达玛

作为中国人民的朋友，1936 年春，阿达玛曾接受熊庆来教授的邀请在清华大学讲学三个月。1963 年 10 月，阿达玛以 98 岁高龄离开人世。《法国数学学会会刊》发布了消息，并配以一幅精美的阿达玛肖像。著名数学家维纳讲过的一个故事可以通过这一肖像得以证实。1935 年，维纳教授在中国讲学时，在北京古玩店淘宝时发现一张中国人的肖像非常像阿达玛。维纳把画买下，送给阿达玛。阿达玛非常喜欢，但后来画还是丢了，或者是让不欣赏这幅画像的阿达玛太太藏起来了。

阿达玛、普桑的证明使高斯、勒让德的猜想成为确定事实。从问题的提出到获得解决，花费了人类一个世纪的时间。这一美妙深刻的结果现被称为素数定理。这一名称就可体现出它的重要性。事实上，素数定理可看作素数分布理论的中心定理。这一结论也被称作全部数学中最令人吃惊的成果之一。一方面，素数虽然看上去是随机的，但当俯瞰整个自然数序列时，却会发现一个十分明显的模式：x 越大，密度越接

近 $\frac{1}{\ln x}$ 。另一方面，这个定理“将素数个数同微积分中与生物增长有关的函数连接在一起”，“刻画了在素数和自然对数函数之间的简单关系，而自然对数一眼看来与素数无关”。

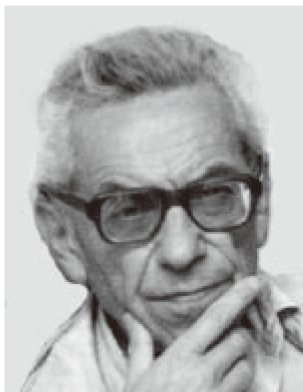
此外，素数定理之所以重要，不仅因为它做了一个关于素数的基本、优雅的叙述，并有超出数学之外的应用，还因为很多新的数学公式都是在证明它的过程中创造出来的。事实上，1896 年阿达玛与普桑的证明过程都是由下一节我们要介绍的黎曼所开辟的，而其证明过程中都运用了高深的解析工具，即复变函数方法。因而，他们对素数定理的证明促进了解析数论的进一步发展。

素数定理的初等证明

1896 年以后，许多数学家认为素数定理不可能找到至多只用一些初等微积分的“初等证明”。1921 年，英国数论大师哈代在哥本哈根数学学会发表演讲时说，虽然“断言一个定理肯定不能用某种方法证明是轻率的”，但是素数定理的初等证明照他看来是不可能的。“如果谁给出了素数定理的初等证明，那他就证明了现在关于数论、解析函数论中何谓深刻、何谓肤浅的见解是错误的……从而就到了该丢掉一些著作来重写理论的时候了。”哈代死于 1949 年，就在他去世后几个月，埃尔德什与他的一个同事塞尔伯格给出了这一定理的相当长却初等的证明。证明中除 e^x 与 $\ln x$ 外，不用任何“超越性”的东西，也不需要高深的微积分。下面我们来介绍一下埃尔德什与塞尔伯格这两位 20 世纪著名的数学家。

埃尔德什

埃尔德什（1913—1996），是匈牙利数学家。



埃尔德什

埃尔德什也属于那种极具数学天赋的数学天才，3 岁时他已能解算 3 位数的乘法，4 岁时独自发现了负数。在大学一年级时，他发表了一篇论文，给贝特兰猜想一个初等证明。上面已经提到，此结论最初由切比雪夫证明，不过他的方法比较复杂。埃尔德什相信有一本超穷的天书，这本天书中有所有数学定理的最漂亮、最完美的证明。而切比雪夫的复杂证明不可能是“天书”的手笔。对比切比雪夫与埃尔德什的两种证明方法，有人评价说：“同样是移栽一枝蔷薇，切比雪夫用的是蒸汽铲车，而埃尔德什只用汤勺就做到了。”埃尔德什的这一胜利通过下面的一支小调传遍了英语国家：

> 切比雪夫说过，我再说一遍
>
> 在 n 和 $2n$ 之间总有一个素数存在

大学期间，埃尔德什还给出贝特兰猜想一种推广结果的一个初等证明，并作为自己的博士论文于 1934 年发表。这篇论文引起了柏林大学著名数学家舒尔的注意。随后，埃尔德什又解决了几个问题。舒尔大为惊讶，称他为“布达佩斯的魔术师”。埃尔德什在崭露头角后，开始了他漫长的数学生涯。

作为 20 世纪伟大的数学家，埃尔德什可谓典型的数学苦行僧，他抛弃了一切物质享受，没有妻子和孩子，甚至居无定所。他的生活安排就是能把他所拥有的时间最大限度地用于数学。在他看来，私有财产就是累赘，他唯一牵挂的财产是他的数学笔记本。他一生写满了 10 本数学笔记，并总是随身带着一本，以便随时记下他的数学灵感。他的思维能力无与伦比，却对日常生活束手无策。他把一生都奉献给了一项专一的事业，那就是发现数学真理。他强调说，生活的目的就在于证明和猜想。甚至在他生命的最后 25 年中，他仍每天工作 19 个小时。当朋友们劝他悠着点时，他总是回答：“坟墓里有的是休息时间。”1996 年 9 月 20 日，他以 83 岁高龄去世。他比历史上任何其他数学家都思考了更多的问题。在论著数目方面，他可与欧拉比肩。据统计，在不同的数学领域内，他共发表了 1525 篇高水平的学术论文，其中 1100 篇论文至少有一位合作者。

事实上，埃尔德什是有史以来合作者最多的数学家，由此在数学界还诞生了有趣的埃尔德什数。对于那些直接与埃尔德什合作的人，人们称其“埃尔德什数”（简称埃数）为 1。这是一个数学界都羡慕的代码，意味着与大师本人合作过。这个“埃数”为 1 的俱乐部成员多达 485 人。这个代码可以依次推下去。“埃数”为 2，意味着与和他合作写过论文的人合作过。“埃数”为 3 是说与埃尔德什的合作者的合作者合作写过论文。已知的仍活跃的数学家的最高“埃数”为 7。像我们这种与埃尔德什挨不上边的，也可以获得一个“埃数”：∞（无穷大）。

埃尔德什又是 20 世纪旅行最多的数学家。他没有固定住所，是一位巡回访问学者，很多时间花在路上。在 60 多年的数学生涯中，他带着两件旧行囊，不停奔波在

各大洲的大学数学系和研究中心之间。他经常会出现在一个数学同事的门阶上，宣布“我的大脑敞开了”，然后就和这位数学家工作一两天，直到厌烦了或者他的东道主疲惫不堪。这时，他又会去拜访另一家。他在 25 个以上的国家研究过数学。他的座右铭是“另一个屋檐，另一个证明”。

埃尔德什在提问题并促使这些问题得以解决方面也堪称世界第一。埃尔德什还以在全世界发掘神童为自己的使命。埃尔德什是这样培养年轻人的：把一个个问题连珠炮似地砸向他们，好像他们是职业数学家似的。埃尔德什在这方面的付出得到了极大回报。许多有数学天赋的年轻人在他的培养下走上了数学研究的道路。

1984 年，他由于在众多数学领域中的贡献，以及他个人与全世界数学家的合作而荣获数学界最负盛名的沃尔夫奖。

埃尔德什的主要贡献在数论、组合学、概率论、集合论和数学分析等方面，他被看作有史以来最伟大的离散数学家。在数论方面，他的工作尤为出色。而素数又是他的重中之重。素数是埃尔德什的亲密朋友，他在素数方面的渊博知识无人可比。常常沉浸在素数世界中的他与素数之间似乎有着某种奇妙的密切关系。对于素数问题，他的洞察力异常敏锐，每当听到有关素数的一个新问题时，他经常很快就超过那些为此耗费了大量时间的人。

在素数问题上，埃尔德什取得的最伟大的胜利就是，在 1949 年用初等方法证明了素数定理。同时获得这一突破成果的还有塞尔伯格。

独行侠塞尔伯格

塞尔伯格（1917—2007），是挪威裔美籍数学家。



塞尔伯格

塞尔伯格的父亲和两个兄弟都是数学教授。受家庭环境的感染和熏陶，他自幼就喜欢数学。大约在13岁时，他开始自学高等数学。当看到莱布尼茨公式后，他感到非常惊奇。在他看来，“$\frac{\pi}{4}=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\cdots$ 这个公式实在美极了；单数1、3、5…这样的组合就可以给出π。对于一个数学家来说，此公式正如一幅美丽图画或风景”。从此，他对数学更加心驰神往。后来他又读了哥哥借的印度数学奇才拉马努金的全集，他的想象力被唤起了，他发现了新大陆。

通过自学以及阅读父亲书架上大量的数学书，塞尔伯格在1935年进入大学时已经能够做出原创的工作了。1942年，塞尔伯格完成博士论文“论黎曼ζ函数的零点”，这是黎曼零点研究上取得的一项重大进展。他在第二次世界大战时做出的这一突

破，在战后的 1946 年开始被人所知。普林斯顿高等研究院的著名数学家外尔（1885—1955）在欧洲做长途旅行时，听说了塞尔伯格的漂亮结果，于是邀请他到高等研究院工作。1949 年，塞尔伯格完成素数定理的初等方法证明，轰动了数学界。1950 年，年仅 32 岁的塞尔伯格荣获菲尔兹奖。此后，他又在许多数学分支做出重要贡献。1986 年，这位成就卓著的数学家又荣获沃尔夫数学奖。

在塞尔伯格看来，数学是一种最激动人心的智力活动。他曾感慨地说："我很同情那些没有研究数学的人，我觉得他们失去了一种最激动人心的丰富智力活动的回报。"这或许正是他生命不止、研究不止的动力所在吧。作为数论家中的一位巨匠，塞尔伯格在长期数学生涯中研究过许多经典问题，并为众多领域贡献了重要的证明和技术。

在研究数学的风格上，塞尔伯格与埃尔德什迥然不同。他习惯于单枪匹马地工作。如在第二次世界大战中，他无法得到国外的科学杂志，他与数学界的交流彻底中断了。对此，他认为是一个激励，"这就像在监狱中一样，你与世隔绝，从而可以全神贯注于你的工作，不会被他人的工作影响。从这点来看，我认为这种状态对于做自己的工作还是相当不错的"。这种自足的态度贯穿了塞尔伯格整个数学人生。这个文静甚至有些孤僻的人，几乎从不与他人合写论文，他唯一的一篇合作论文是与印度数论家周拉合作完成的。另一次差一点完成的合作恰发生在他与埃尔德什之间，两人曾打算在素数定理的证明方面合写一篇论文，然而最终悲剧性的结局却给两人都留下了深深的阴影。关于两人在素数定理证明上的争端曾在几十年中一直是数学界的传闻和猜测的一个热点。事情的原委大致如下。

1948 年 3 月，善于用初等方法攻击数论难题的塞尔伯格发现了现在所称的"塞尔伯格公式"。当他认识并喜欢的匈牙利数学家图兰准备离开普林斯顿时，塞尔伯格给图兰看了自己的这一公式，但"没有告诉他这个公式的证明，也未告诉他这一公式的可能推论及我在这方面的想法"。事实上，塞尔伯格正准备用这一公式得出素数定理的一个初等证明。在告诉图兰自己的公式，但对自己的下一步打算只字未提后，塞尔伯格

离开了研究所 9 天。就在他离开期间，图兰在研究所做了一次演讲，演讲结束后的简短讨论中，图兰提到了塞尔伯格公式。在座的图兰的好朋友埃尔德什立刻意识到塞尔伯格公式可能与素数定理关联。他的直觉告诉他，可以先利用塞尔伯格公式推出一个结论，即当素数增大时，两相邻素数之比趋于 1。如果这一中间定理得到证明，他就可以进而去实现素数定理初等证明的最终目标。当塞尔伯格回来时，发现埃尔德什已经满怀热情地投身于这一工作，并且不久埃尔德什就告诉了他自己打算证明的中间定理。深感不悦的塞尔伯格为了给埃尔德什泼冷水，就说他怀疑无法从自己的公式导出埃尔德什要证明的中间定理，因此很可能无法导出素数定理的初等证明。塞尔伯格甚至还布了一个迷魂阵，说他已构造出一个反例。这一反例意味着，从塞尔伯格的公式不能推出素数定理。这种想把埃尔德什引入歧途的做法不仅没有成功，而且为后来的争论埋下了祸根。第二天，埃尔德什就告诉塞尔伯格，说他已证明了中间定理。很快，塞尔伯格利用埃尔德什的定理完成了素数定理的证明。埃尔德什非常高兴，并设想两人一起发表一篇论文，并讲明各自对这一证明的贡献。

论文尚未写成，埃尔德什已开始像以前习惯的那样，向他众多的通信者寄明信片报喜，告知他们说他已与塞尔伯格攻克了素数定理。结果，当时远不如埃尔德什有名的塞尔伯格所碰到的每一个人都将这一证明“完全地或至少是实质上”归功于埃尔德什。一件偶然事件让塞尔伯格感到丢脸。一位陌生的数学家在碰到他时，对他说：“你知道吗？埃尔德什和那个谁已经用初等方法证明了素数定理了。”被深深激怒的塞尔伯格立刻坐下来，重新投入研究。结果他找到一条通过他的公式推出素数定理的新途径，这一方法不需要依赖埃尔德什的贡献。于是，他决定单独发表自己的证明。

把塞尔伯格的误导当真的埃尔德什相信，他在这一问题的解决中起到了关键作用，因此他对塞尔伯格的做法非常不满。最终，塞尔伯格将其论文投给有声望的《数学纪事》。埃尔德什的文章先投到《美国数学学会画报》，但被拒绝发表，后发表在《全国科学院学报》上。1950 年，塞尔伯格获菲尔兹奖，其中一个主要原因是他对素数定理

的初等证明。1952 年，埃尔德什也因为给出素数定理的初等证明，获得了与菲尔兹奖荣誉相近的柯尔奖。两人的贡献都赢得了数学界的认可。

事后分析这起事件，我们可以说，两人不同的研究方式是导致这次冲突的一个重要原因。对埃尔德什来说，数学的目的在于证明与猜想，并且要尽可能快地去证明与猜想。为此，他乐于寻求合作者，并与别人分享发现。在这种共享中，埃尔德什常表现出罕有的坦荡襟怀。但对塞尔伯格来说，他宁愿按自己的步调孤军奋战，埃尔德什进攻性的数学研究方式太过出格，令人难以接受。不同的数学家有不同个性，也有不同的研究方式，这是完全可以理解并应表示尊重的。就历史而言，20 世纪以前，数学的合作可以说是凤毛麟角，那时的重大结果大多只与一个人的名字有关。但 20 世纪以后，数学研究方式逐渐发生了重大转变，合作发表论文不仅不足为奇，而且已形成新的风气。在这种情况下，合作者之间如何处理好双方关系就成为一个值得玩味的问题。

塞尔伯格、埃尔德什对素数定理的初等证明，是素数研究中的一项重大突破，它表明在对付素数定理这样重要的结论时，并非一定要使用“飞机、大炮”，用“步枪”就够了。让数学界大吃一惊的这一证明在当时引起了轰动效应。然而这个很美的证明并没有产生如哈代所预料的革命性效果，没有哪本书因此要被抛弃，也没有一种理论因此而需要重写。回顾一下，这并非数学中如此重要的作品。事实上，素数定理的初等证明已是一颗昨夜星辰，现在它已差不多成为一个脚注了。

第三节
素数的音乐与黎曼零点

高斯、勒让德的工作标志着素数研究方向的转变。在此之前，多少年来，人们总是倾向于去预测下一个素数的准确位置，或是找到某个可以生成素数的公式。然而，这类努力全以失败而告终。而高斯与勒让德不再拘泥于预测下一个素数的确切所在，不再关注究竟哪些数是素数，而是转向一个全新的问题：任意给定一个数x，是否存在某个方法，可以估计从1到x之间素数的个数？通过退后一步并问了这样一个更普遍的问题，令人惊讶的规律从迷雾中显现了。

换言之，在此之前，人们在倾听素数的音乐时，注重的只是单个音符，无法听到整个作品，然而当考虑究竟存在多少个素数时，人们才找到了倾听主旋律的方法。多年后，德国著名数学家黎曼在此基础上，把对素数分布的研究推向壮丽的巅峰，奏响了一曲有关素数分布的神秘乐章，并为后世数学家留下一个魅力无穷的伟大谜团。

黎曼与8页论文

黎曼（1826—1866），是历史上最具想象力与独创力的数学家之一。



黎曼

1826 年 9 月 17 日，黎曼出生在德国的一个小村庄。从幼年起，他就是一个胆小、缺乏自信的人，害怕在公开场合讲话和引起人们的注意。大约 6 岁时，他的数学才能便显露出来。他不仅解决了所有留给他的数学问题，而且还编造更困难的难题来捉弄他的兄弟姐妹。10 岁时，他跟着一位职业教师学习更高级的算术和几何，而他很快超过了老师，并且往往得出更好的解答。在中学时，学校的校长注意到他的数学才能，让他随意使用自己的图书，并允许他可以不上数学课而自学数学。16 岁时，黎曼借走了校长书架上的勒让德的《数论》，这是一本 859 页大四开本的书。6 天后，黎曼送回了书，当校长问他“你读了多少”时，黎曼回答说：“这本书写得非常奇妙，我已经掌握了它。”在两年之后的毕业考试中，校长拿勒让德书里的一些问题来考他，结果黎曼的成绩非常好。也多亏勒让德，年轻的黎曼在心中种下了一粒种子，在未来的日子里它将会绽放出绚丽的花朵。在中学时，黎曼以惊人的速度进行自学，他不仅仅学习了

勒让德的著作，还学习了欧拉的著作，熟悉了微积分及其分支。如这位校长后来回忆的，“在当时，他就已经可以说是一位数学家了。站在他的身旁，作为老师的我感到自己知识的贫乏”。

1846 年，19 岁的黎曼进入哥廷根大学。一年后，他转入柏林大学，成为雅可比、狄里克雷等杰出数学家的学生。

黎曼的兴趣是广泛的，他对哲学也很感兴趣，他用于物理的时间与用于数学的时间不相上下。作为一个物理数学家，他对数学中很可能具有科学应用价值的东西的直觉，与牛顿、高斯和爱因斯坦是同等的。“要是他多活二三十年，很可能会成为 19 世纪的牛顿或爱因斯坦。”大学期间，由于对物理学的着迷，1850 年他参与了一个数学物理研究班。由于耗费了过多的时间，他的数学博士论文直到 25 岁时才交出。

这篇晚交的论文题目是《单复变量函数一般理论基础》，于 1851 年 11 月初，呈交给高斯审查。在写给哥廷根大学的正式报告中，高斯发表了对黎曼论文的正式看法：“黎曼先生交来的论文充分说明了，作者对所论述的问题做了全面深入的研究，作者具备真正有创造性且活跃的数学头脑。作者的表达方式清晰简明，一些地方很优美。……它远远超出了博士论文所要求的标准。”

一个月后，他完成了论文答辩。他开始希望得到一个讲师职位，但还需要作一次就职演说。在第三章中，我们已经介绍过，黎曼这次著名的演讲开辟了以后被称为黎曼几何的这一新的数学分支，为后来的几何化物理铺平了道路，其工作后来成为广义相对论的数学基础。

1857 年，31 岁的黎曼得到助理教授的位置。33 岁时，黎曼接替了去世的狄里克雷成为正教授，经济状态由此有了改善。然而，不幸的是，黎曼生来体质很弱，在未能获取他十分之一金色收成之前，他于 1866 年 7 月 20 日因病离开了这个世界，年仅 39 岁。“他像流星一样出现然后消失，他活跃的时间只不过 15 年”，留下的全部数学

成果只有八开本的一卷。然而其为数不多的著作都洋溢着新奇的思想和多产的概念，指引着别人去开展大量重要的研究。“尽管黎曼的文章只有538页的一卷，但如果用理性来衡量，这一卷重如千钧。他众多发现中的每一个都将注定改变数学的进程。”

在他为数不多却异常深刻的数学著作中，我们需要提到的是他在数论方面的贡献。早在阅读勒让德的书时，黎曼就对素数之谜产生了兴趣。勒让德在书中给出的估算素数个数的经验公式更激发了黎曼的灵感。为了改进勒让德的结果，黎曼写出并公开发表了自己在数论方面的唯一一篇论文。这是一篇不到10页但内容极其深刻，并极具启发性的论文。这篇论文的题目是《论小于给定数的素数个数》，于1859年11月发表在德国科学院院刊上。黎曼的这一论文提供了研究数论的新思想和新方法，开辟了新的研究数论的方向。事实上，这篇文章的发表翻开了解析数论研究的新篇章，从此微积分的有力方法被广泛应用到自然数问题的研究上。

为了领略黎曼的思想，我们需要先从欧拉的“数学上最引人注目的发现之一”谈起。

在1737年左右，欧拉引进了一个特殊函数 $\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\frac{1}{4^s}+\cdots$，同时他发现这一 ζ 函数与一个只有素数的级数之间存在奇妙关系，即 $\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_{p}\frac{p^s}{p^s-1}$，其中 $\prod$ 代表连乘积。公式中左边的求和对所有的自然数进行，右边的连乘积只对所有的素数进行。举一个例子，取 $s=2$，则欧拉的发现是，$\zeta(2)=\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\prod_{p}\frac{p^2}{p^2-1}=\frac{4}{3}\times\frac{9}{8}\times\frac{25}{24}\times\cdots$。

欧拉发现的这一无穷乘积公式是他素数研究中所做出的最为重要的贡献之一，曾标志着解析数论的开始。所谓解析数论，是数论中用分析方法研究数的性质的一门分支。用解析数论研究素数模式的关键在于找到能提供素数信息的函数。而欧拉 ζ 函数正是人们寻找到的第一个这样的函数。利用它，可以将自然数级数翻译成另一素数级数。于是自然数与素数具有的这种关系，使欧拉 ζ 函数成为数学家理解素数的一种有

效工具。欧拉本人曾用这一工具给出了素数无穷多的另一有趣证明。1837 年，德国数学家狄里克雷对欧拉 ζ 函数做了推广，提出狄里克雷 L 函数，并由此证明了著名的等差数列的素数定理。可以说，狄里克雷的工作进一步奠定了解析数论的基础。

沿着欧拉开辟的道路，向前迈出更重要一步的正是黎曼。

欧拉 ζ 函数，作为一个实数（定义在实数上）到实数（值为实数）的函数，是一个一维对象。因此，虽然通过欧拉的无穷乘积可以让它与素数相联系，但它没有丰富的几何结构来帮助人们揭示素数的模式。黎曼所做出的第一步是，用复数 z 代替实数 s，使 $\zeta(z)$ 的值也成为复数，即取 $\zeta(z)=\sum_{n=1}^{\infty}\frac{1}{n^z}$（$z=a+bi$）。但这种推广存在一个问题，这一无穷和对某些复数没有意义。黎曼为此迈出了第二步，他用一种称为解析延拓的精巧数学方法解决了这个难题。在这种处理后，除了 $z=1$ 这一特例之外，我们可以计算任何复数 z 对应的 $\zeta(z)$ 值。经过这样解析延拓后得到的函数，仍记作 $\zeta(z)$。

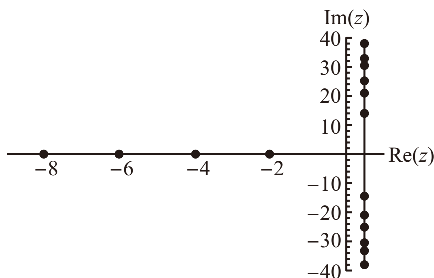
在历史上，黎曼并不是第一个在欧拉 ζ 函数中用复数 z 代替实数 s，从而迈出第一步的人，但他却是迈出第二步的第一人。更重要的是，黎曼对这一函数做了深入研究，因此复变函数 $\zeta(z)=\sum_{n=1}^{\infty}\frac{1}{n^z}$ 现在被恰当地称为黎曼 ζ 函数。

下面我们就来简单看一下，黎曼在 1859 年的著名论文中是如何运用 ζ 函数探讨素数模式的。

在上一节中，我们介绍过高斯给出了 $\pi(x)$ 的一个很好的估算公式 $\mathrm{Li}(x)$。那么，高斯的结果是否还可以进一步改进呢？黎曼对此做了回答。他先是给出一个新公式，我们不妨记其为 $R(n)$，这一公式与高斯的结果相比更为精确。比如，在小于 1 亿的数中，高斯的对数积分公式多预测了 754 个素数，而黎曼的公式仅仅多出 97 个，这差不多是十万分之一的误差。这个公式从根本上改进了高斯的估计。但黎曼仍不满足，他不希望有任何误差。而他也确实找到了这个计算素数个数的完全精确的表达式。这一可以完全消除误差的方法正与他所引入的黎曼 ζ 函数紧密相关，更具体而言，是与黎曼

ζ函数的零点紧密相关。

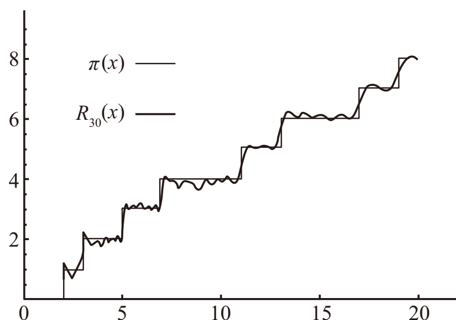
所谓黎曼ζ函数的零点，就是使$\zeta(z)=\sum_{n=1}^{\infty}\frac{1}{n^z}=0$（$z=a+bi$）的$z$值。黎曼指出，$\zeta$函数有一些相当明显的零点$z=-2n$（$n$为正整数）。这些性质简单的零点被称为黎曼$\zeta$函数的平凡零点（trivial zero）。除这些零点外，黎曼ζ函数还有许多其他零点，这些比平凡零点更复杂的零点称为非平凡零点（non-trivial zero）。对数学家而言，真正重要的正是这些非平凡零点（在不引起误解的情况下，后面有时会把这些非平凡零点直接称为零点或黎曼零点），因为正是这些零点蕴藏着揭开素数秘密的关键。比如，利用这些非平凡零点的坐标，可以得到准确的素数个数，即给出$\pi(x)$准确的值。对此，我们只能简略地描述一下。



在引入黎曼ζ函数后，黎曼发现利用ζ函数，每个非平凡零点都能转化为其自身独特的波。这样的零点有无数个，因此存在无穷个波，这个点越偏北，即离实轴（上图中的Re(z)轴）越远，对应的波高越大。在$R(n)$的估计基础上，加上数n处的每个波高，不同的波高正好可抵消素数个数估计中的误差，从而可得到数n以内素数个数的准确公式。这个公式可以用文字简单地记为“素数 = 零点 = 波”。但注意，并非每个素数都对应一个零点。如果我们将这个波看作声波，那么可以得到更为生动的类比：每个非平凡零点都像是由音叉发出的、不含任何和弦的纯音符。每个这样的零点都发出一个音符，而零点的南北坐标决定了对应声音的音高（也就是波的频率）。而当这些基本的音符被同时奏响时，产生的就是素数的音乐。因此，素数曲线就像管弦乐队的声音，而零点就像是把该声音分解成不同的频率，使得每个零点对所有素数都产生

作用。另一个帮助理解的类比是，把黎曼零点的整个集合看作是素数的全息图。如果能够“看遍”所有的零点，那么我们就会看到素数的精确图像，并能发现想要知道的一切。

借助于图形我们可以更好地看出，如何通过在黎曼素数个数公式上加上修正波来改进这个公式的效果。在下图中，呈阶梯状图形描述的是素数个数。而呈锯齿状图形表示的是，黎曼函数 $R(n)$ 的图形与前 30 个零点（所谓“前”意味着靠近实轴）产生的波的合成效果。



当黎曼 ζ 函数图像向北前进时，每碰到一个零点我们就加上其相应的波，而这个波将会对原来的图形进行微小扭曲。这样加上的波越多，得到的合成图像就与素数个数的阶梯状图像越接近。加上无穷多个零点产生的波后，最终的图像将与素数个数的阶梯状图像完全吻合。

黎曼的发现，揭示出素数与零点之间的紧密联系。于是，对零点分布进行探讨成为极其必要的事情。在“最杰出和具有最丰富内容”的 8 页论文中，黎曼写下了自己对此高屋建瓴般的思考，并提出了关于非平凡零点分布的几个猜想。其中最重要的一个猜想被称为黎曼猜想，或称黎曼假设，其内容很简单：黎曼 ζ 函数的所有非平凡零点都具有形式 $\frac{1}{2}+b\mathrm{i}$ 。换句话说，黎曼 ζ 函数的所有非平凡零点都位于复平面上实部为 $\frac{1}{2}$ 的直线上。这条直线后来被称为临界线。

于是我们看到，在黎曼的引导下，一条素数研究的前所未有的道路被开启了。先是通过黎曼 ζ 函数，随机的素数与 ζ 函数的非平凡零点连接起来。素数的分布问题转化为黎曼零点的分布问题。而零点的分布又归结为黎曼猜想。由此，黎曼猜想在素数研究中占据了中心地位。在黎曼卓越的思想被领悟后，懂得这一问题价值的人们开始以巨大的热情去攀登这座数学界的珠穆朗玛峰。下面我们就来介绍一下，在通向揭示黎曼猜想真假的坎坷道路上探索者们留下的足迹。

数学接力棒

黎曼迈出了关于黎曼 ζ 函数非平凡零点分布的第一步。在其论文中，他断言，黎曼 ζ 函数 $\zeta(z)=\sum_{n=1}^{\infty}\frac{1}{n^z}$（$z=a+b\mathrm{i}$），所有非平凡零点都位于复平面上的带状区域 $0\leqslant a\leqslant 1$ 内，即所有非平凡零点都落在实部属于 [0, 1] 的带状区域。黎曼的这一断言，可以很容易被证明。

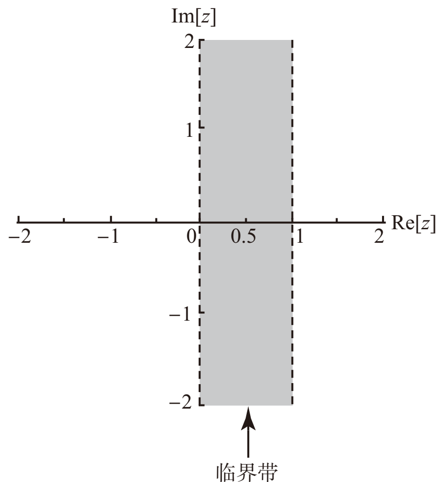
在对零点分布的研究中，黎曼还指出黎曼非平凡零点具有对称性。首先，非平凡零点关于东西向的实轴是对称的，即上半复平面与下半复平面的非平凡零点是一一对应的。由这一对称性，讨论时可以只考虑虚部（即 $\mathrm{Im}(z)$）大于零的零点。人们又把这些零点以虚部大小为序排列，这样“前 N 个零点”指的是虚部最小的 N 个零点。另外，临界线（即穿过实部坐标 $x=\frac{1}{2}$ 的南北向直线）也是一条对称轴。

以上两点表明：研究零点的分布，只需研究带状区域 $0\leqslant a\leqslant 1$ 的右上部分，即 $\frac{1}{2}\leqslant \mathrm{Re}(z)=a\leqslant 1,\ \mathrm{Im}(z)=b>0$ 就可以了。

在黎曼论文发表30多年后，关于黎曼猜想的研究出现了第一个重要进展，而这一进展正是在证明素数定理时得到的。

1895年，有数学家证明了：素数定理等价于黎曼ζ函数在$\mathrm{Re}(z)=1$的直线上没有非平凡零点。这为素数定理的最终证明铺平了道路。1896年，阿达玛与普桑两人独立证明了：直线$\mathrm{Re}(z)=1$上没有非平凡零点。由此数学界期盼了世纪之久的素数定理获得了证明。而这一与黎曼猜想相关的证明也成为黎曼猜想上的一个突破。

这一工作使人们对黎曼ζ函数非平凡零点分布的信息稍近了一步。由于黎曼零点分布的对称性，我们可以进而知道，在直线$\mathrm{Re}(z)=0$上也没有非平凡零点。由此，黎曼得到的"所有非平凡零点都位于复平面上的带状区域$0\leqslant a\leqslant 1$内"的结果可以改进为"所有非平凡零点都位于复平面上的带状区域$0<a<1$内"。后来，这一区域被称为临界带（critical strip）。



显然，这一结果与黎曼猜想还相距很远，但它对数学的发展带来了重要影响。一方面，素数定理的证明是利用解析函数论在黎曼ζ函数上取得突破而得到的。这样数学分析的威力得到数论专家的进一步承认，从此解析数论得到迅速发展。另一方面，由这一远远弱于黎曼猜想的结果即可推出素数定理，由此可一窥黎曼猜想被证明后的

影响。因此，素数定理的证明，大大提高了数学界对黎曼猜想的关注度。

1900 年，伟大的德国数学家希尔伯特在巴黎第 2 届国际数学家大会上做了题为“数学问题”的有划时代意义的演讲。演讲中列出了 23 个数学问题，其中第八个问题是“素数问题”，黎曼猜想被列入其中。从此，黎曼猜想成为令整个数学界所瞩目的难题。一个在数学界流传的故事可以说明这一猜想在希尔伯特心中的地位。

德国流传着一个关于腓特烈大帝一世巴巴罗萨的传说。德国百姓喜欢巴巴罗萨，他死于一次十字军东征，死后埋在遥远的墓穴里，但谣传他还活着，沉睡在基法塞尔山的一个洞里。只要德国需要他，即便千百年后他也会苏醒过来。有人因此问希尔伯特：“假如你能像巴巴罗萨那样复活，那么 500 年后你会做什么？”“我想问，”希尔伯特说，“有谁证明了黎曼假设了吗？”

随着时间的推移，黎曼猜想越来越成为数学家心目中的“圣杯”。然而，赢得这一圣杯的道路却异常曲折。

在阿达玛、普桑之后，对黎曼猜想发动成功攻击的是朗道与哈拉德·玻尔。

哈拉德·玻尔，丹麦著名数学家，伟大物理学家尼尔斯·玻尔的弟弟。有些与众不同的是，他还是一名出色的足球运动员，当获得数学博士学位时，报纸的体育版对此做了报道，上面的照片是他手拿一个足球。作为丹麦足球队的主力，他曾参加 1908 年奥运会并获银牌。作为数学家，他在哥本哈根工作，但同许多数学朝圣者一样，他会“打起背包”前往当时的世界数学中心哥廷根大学。在那里，他与著名数学家朗道进行了合作。

朗道就是我们上一章提到过的那位负责审查费马大定理证明并想出巧妙办法的人。这个并不容易相处的人对学生非常严格。一则轶事说，有一次，他的一个学生结婚后准备去度蜜月，火车刚要离开车站时，朗道从天而降，从窗口塞进一叠新书手稿，并命令道：“校对完毕，回来交给我！”

1914年，玻尔与朗道在合作中考虑了$\left[\frac{1}{2},\frac{1}{2}+\delta\right]$（$\delta>0$）这一区域内的零点个数，并与在此范围之外的零点数相比较，他们证明无论δ多么小，在上述带状区域内的零点都占所有零点的绝大部分。也就是说，包含临界线的无论宽度多么窄的带状区域都包含了几乎所有的非平凡零点。其结果用更简单的话表述就是，几乎所有非平凡零点都落在临界带内，或者说几乎所有非平凡零点都聚集在临界线附近。然而，值得一提的是，这并不意味着大部分零点落在临界线上。实际上，这一结果对究竟有多少零点落在临界线上没有提供任何信息。也就是说，从朗道与玻尔的这一结果我们无法得知究竟有多少零点落在临界线上。

在黎曼假设问题上，朗道还做出了另一贡献，在其《素数分布理论手册》（1909）一书中，他介绍了素数与黎曼函数之间的奇妙联系。在这本书出版之前，黎曼的素数故事在数学界中只是小范围流传。而“这本书的出现让一门仅有数位探索者的学科，变为近30年来最热门的学科之一”。正是受这本书的影响，同一年英国数学家哈代也在探索黎曼假设的征程中迈出了重要一步。

哈代（1877—1947），童年时代就显示出其数学才能。13岁时，他获得奖学金进入当时以数学家摇篮而著称的温切斯特学院学习。1896年，他又获得入学奖学金并进入剑桥大学三一学院继续深造。在校期间发生了两个小插曲。一个是宗教方面的，哈代决定不相信上帝。对他来说，上帝是他的敌人。第二个则与考试有关。当时，剑桥有一种数学荣誉学位考试，考试优胜者可以获得极大的荣誉。然而，在哈代看来，这种非常呆板的考试没有意义，因为它根本不利于发挥学生的数学想象力、洞察力或作为一个有创造性的数学家所应具备的其他素质。但他不得不屈服于这种考试制度。1898年，他在数学学位考试中获得第四名。1900年，在更权威的数学荣誉学位第二部分考试中，他荣获第一名，并得到了三一研究基金。为此，他不得不经受痛苦煎熬。好在，哈代从一位应用数学家拉弗那里收获了幸运。“第一个使我拨云见日的是拉弗教授，他教了我几个学期，使我第一次理解了分析这个概念。但最使我感激的是他建议我阅读

若当的名著《分析教程》。我永远不会忘记我读那本杰作时的震惊，这是我这代数学家受到的第一个启迪，读这本书时我才第一次认识到数学真正意味着什么。也是从那时起，我走上了一个真正的数学家的道路，对数学树立了正确的目标，产生了真正的激情。”于是，在结束令他难以忍受的考试之后，他便以极大的热情投入到真正的数学研究中，并于此后的岁月里在多个数学领域做出了重要贡献。



哈代

哈代在黎曼假设上得到的突破性结果是，有无穷多个零点落在黎曼临界线上。要知道，在这一结果被证明之前，人们通过计算找到的具体零点还少得可怜，而在理论研究中，还没能证明哪怕一个零点落在临界线上。因此，哈代的证明一经发表，即令欧洲大陆数学界为之震惊。在这次成功的尝试后，黎曼假设成了哈代最希望攻克的难题，它占据了哈代相当多的时间。在一份新年决心清单上，他列下的第 1 个目标就是“证明黎曼假设”，还有另外 5 个不可能实现的目标，如第 6 个目标是“刺杀墨索里尼”。相比之下，黎曼假设的证明看上去更容易一些。

除热衷数学研究之外，哈代还对板球非常感兴趣。当每年的板球赛季于 9 月结束后，喜欢阳光的哈代总是习惯在英国大学新学期开学之前去拜访哥本哈根的玻尔。两人每天的工作很程式化：早晨放一张纸在桌子上，哈代在上面写下当天将要完成的工

作内容——证明黎曼猜想。但他们一次次地失败。

一次，因为新学期的到来，哈代不得不返回英国。当他准备踏上小船时，正赶上当时海面上波涛汹涌，哈代很担心自己的安危。不过，他还是上了船，但给玻尔寄了一张明信片："已经得到了黎曼假设的证明，由于明信片太小而无法写下哈代。"在哈代看来，他的敌人的上帝不会让小船沉没，因为如果船沉了，哈代淹没了，世人就会认为他与黎曼假设的证明一起葬身海底了。上帝当然不会让他享有这么大的荣誉。结果哈代安全地返回了英国，而黎曼假设的证明仍悬而未决。

与埃尔德什一样，哈代很善于合作。在他的密切合作者中，有与他共同工作了4年的印度传奇数学家拉马努金，而他最著名的合作者则是李特伍德。



李特伍德

李特伍德（1885—1977），比哈代年轻8岁。他也是剑桥大学的学生，而且赢得过数学荣誉学位考试第一名。他与哈代一样，只是把这种考试看作不得不参与并且要获胜的事。考试一结束，他就准备在真正的数学研究中一试身手。他询问导师，希望能找一个恰当的问题来打发漫长的暑假。当时英国数学界很孤立，他的导师对黎曼假设理解不深，他认为"黎曼假设适合于最聪明的学生，而李特伍德就是这样的学生"。于

是，他的导师平静地让他去试证黎曼假设。对这一问题的重要性、难度都一无所知的李特伍德带着极大的兴奋和自信开始了自己的研究。在这次与黎曼猜想的最初接触中，李特伍德并非一无所获，他得到了一些为欧洲大陆数学界所熟知的结果。更大的收获是，从此他开始“痴迷于这个难题”。在他的学术生涯中，他花费很多时间去研究黎曼 ζ 函数和黎曼假设，并在此过程中证明了好几个相关的结果。

其中一项研究前面提到过。在 1914 年，他严格证明了：只要 x 足够大，总会存在一个区域，在此区域中 $\mathrm{Li}(x)$ 给出的素数个数会小于真正的素数个数 $\pi(x)$，从而推翻了高估素数猜想。

另一项直接涉及黎曼假设的重要结果是他与哈代于 1921 年做出的。两人在哈代 7 年前结果的基础上又迈进了一步，他们发现临界线上不仅有无穷多个零点，而且其趋于无穷的速度起码是 KT（其中 K 为大于零的常数）。

这一精彩的结果只是哈代和李特伍德合作结出的硕果之一。两人的合作是一个神奇的动人故事。自 1910 年起直到哈代 1947 年去世，他们进行了长达 37 年的合作。在此期间，他们合写了 100 多篇论文。作为完美的搭配，两人的合作基于如下具体原则。

- 规则一：无论他们写给对方的信是否正确，都不重要。
- 规则二：没有强制性的阅读并回复对方来信的要求。
- 规则三：两人尽量不要同时考虑相同的问题。
- 规则四：这是最重要的原则，为了避免争吵，所有的论文都由两人共同署名，即便其中一人对此毫无贡献。

正如玻尔评价的：“从来没有人基于这样消极的规则，完成了如此重要并且融洽的合作。”对数学史上最著名的这对合作，当时有人曾开玩笑说：“现在只有三个真正伟大的英国数学家——哈代、李特伍德以及哈代－李特伍德。”最后一个指的是两位同样

杰出却有完全不同个性的科学家多年的绝妙合作。这种合作造就了十分重要的成果以及全新的方法，而且远不限于数论这个领域。在人们看来，两人的完美合作使他们几乎融为一体了。

哈代、李特伍德于 1921 年合作得到的定理作为黎曼猜想的最好结果保持了 20 多年。然而如果考察这一结果会发现，它所确立的位于临界线上的零点数目相对于零点总数而言，其渐进比例等于零。因此，它离黎曼猜想的目标还极为遥远。虽然在随后的岁月中，哈代、李特伍德又耗费了无数心血，然而都未能将黎曼猜想的证明再推进一步。事实上，到后来因为证明黎曼猜想的屡次失败使哈代晚年开始倾向于相信：黎曼猜想之所以迟迟无法得到证明，或许只是因为这一猜想是错的。

在哈代、李特伍德的工作之后，接过接力棒在黎曼假设方面取得下一个突破的是我们上一节介绍过的塞尔伯格。

第二次世界大战中，塞尔伯格在与数学界隔绝的情况下，深入研究了哈代与李特伍德的工作。1942 年，他终于在两人研究的基础上，通过改进其结果，成功地向前迈进了一步，并开辟了进军黎曼猜想的新途径。他引入的新想法是，比较临界线上线段 $\{0<t<T,\ x=\frac{1}{2}\}$ 与临界带内矩形 $\{0<t<T,\ 0\leqslant x<1\}$ 中零点的个数，若前者的零点个数记为 $N_0(T)$，后者的零点个数记为 $N(T)$，则黎曼猜想无非是想证明 $N_0(T)=N(T)$。显然，由于上述线段是矩形的一部分，因而 $N_0(T)\leqslant N(T)$，于是在这种思路下，证明黎曼猜想等价于证明 $N_0(T)\geqslant cN(T)$，且 $c=1$。

在 59 页的论文中，塞尔伯格证明了：存在一个常数 A，使得 $N_0(T)\geqslant AT\ln T$。结合以前的研究，他的这一证明相当于，存在一个常数 c，使 $N_0(T)\geqslant cN(T)$。这意味着，确实存在一个比例 c，并且随着考虑更多的零点，这个比率不会缩小为零。但这个比例究竟是多少呢？塞尔伯格的论文中并没有给出具体的数值。但据推算，这个 c 很小（有人认为 c 约为百分之一）。显然，此结果离黎曼猜想 $c=1$ 相距甚远，但他的成果是开拓性的。毕竟在此之前，哈代、李特伍德的研究虽然证明有无穷多的零点落在临界

线上，但未能证明这无穷多个零点是所有零点的一个分数。而塞尔伯格的结果清楚表明临界线上非平凡零点所占比例大于零。因此，其突破被认为是黎曼猜想研究中的一个重要进展，塞尔伯格也因此声名鹊起。

又过了 20 多年，有关这一比例的具体计算终于有了突破性的进展。这一进展是由美国数学家勒文森（1912—1975）做出的。

1974 年，勒文森使用一种新方法得到了对临界线上零点比例的更高估计。最初，他得到一个非常乐观的计算结果：98.6% 的零点位于临界线上！当把自己的一份手稿交给一位同事时，他幽默地宣称自己可以把这个比例提高到 100%，但他要把剩下的 1.4% 留给读者去做。信以为真的这位同事于是开始传播勒文森证明了黎曼猜想的消息。然而，不久后，人们在勒文森的结果中发现了一个错误。在改正后，98.6% 降为 $\frac{1}{3}$。也就是说，他证明了至少有 $\frac{1}{3}$ 的零点位于临界线上。1975 年，勒文森本人在经过异常繁复的计算后，做了一个很小的推进，他把下界估计提高到了 0.3474。令人惊叹的是，勒文森是在年过花甲后获得其重大突破的。不久后，这位在黎曼猜想研究中留下深深印记的数学家死于癌症。

在勒文森之后，许多数学家使用他的方法艰难推进。1980 年中国数学家楼世拓与姚琦把结果推进到 0.35；三年后，这一纪录提高到 0.3685。1989 年，纪录被提高到 0.4。这是迄今为止数学家们在这一方向上获得的最好结果。目前，数学家们普遍认为用勒文森的方法不可能把临界线上零点比例的下界估计推进到 100%。

更不幸的是，即便用这种方法真的把比例提高到 100%，即证明了“100% 的零点位于临界线上”也不等于证明了黎曼猜想，原因在于这里 100% 的意思与我们通常理解的意思有差别。举一个例子，取前 100 个自然数，里面有 75 个合数，于是可以说 75% 的自然数是合数；取前 1000 个自然数，里面有 832 个合数，于是可以说 83.2% 的自然数是合数；取前 100 000 个自然数，里面有 90 408 个合数，于是可以说 90.4% 的自然数是合数……随着所取的自然数越来越多，合数占的比例越来越大。当所取自然

数无穷增大时，合数所占的比例就增大到 100%。但我们很清楚，这并不意味着全部自然数都是合数。换言之，这种渐进意义上的 100% 并不等同于“全部”。

计算零点

通过上面的介绍，我们看到了数学家们在证明黎曼猜想方面所做的努力与缓慢进展。与此同时，数学家们还在计算零点方面艰难跋涉。

计算黎曼 ζ 函数的非平凡零点并非易事。对于一般人而言，恐怕连表达式 $\frac{1}{n^{a+bi}}$ 会有什么数值上的意义都难以理解。事实上，如果想弄懂数学家如何求一个数的虚数次幂（比如求 7^{3+12i}），那么我们需要去复分析和数论班学习几个月。对于数学家而言，寻找具体的非平凡零点之路也异常艰辛。

1903 年，在黎曼的论文发表 44 年后，一位丹麦数学家使用 18 世纪 30 年代发展起来的欧拉－马克劳林方法，在付出了几年艰辛的努力后，首次得到了前 15 个零点，如 $\frac{1}{2}+14.135i$，$\frac{1}{2}+21.022i$，$\frac{1}{2}+25.011i$，…这是人们第一次看到实实在在的非平凡零点。1914 年，人们得到了前 79 个零点。1925 年，在使用了哈代与李特伍德改进后的计算方法后，人们得到了前 138 个零点。进展如此缓慢的原因一方面在于当时还只能用手算，另一方面在于当时使用的计算办法（包括哈代的改进方法）太过复杂。不久后，数学家西格尔的意外发现终于改变了数学家在计算零点时举步维艰的困境。

西格尔（1896—1981），是德国著名数学家。当他还是一个害羞的孩子时，便与朗

道的父亲偶然相识了。朗道的父亲把儿子写的两卷素数书赠给西格尔，并说现在这本书对他而言太难了，但以后就能看懂了。朗道的书被西格尔视若珍宝，并在他后来的数学研究中产生了持续不断的影响。



西格尔

1915 年西格尔中学毕业时，第一次世界大战正激烈进行。对战争很反感的他在进入柏林大学后，选择了与人间世事最不相干的天文学作为自己的专业。然而，他入学那年赶上天文课程开得比较晚，于是他去听了一位数学家的数论课。这一偶然的情况改变了他的人生旅途，最终把他引向数论的殿堂。

1917 年，他被征召入伍，但他拒服兵役，结果被送进一家精神病院。朗道的父亲设法把他救了出来，后来西格尔说，如果不是朗道父亲的话，他很可能已经死了。1919 年，这位年轻人来到哥廷根，见到了自己的偶像朗道，从此他的数学才能开始得到充分发挥。1978 年，西格尔荣获首届沃尔夫数学奖。

在对黎曼猜想的研究中，西格尔留下了许多印记，其中最重要的是他重新发现了黎曼的秘密公式。事情要从黎曼去世后说起。

1866 年，年仅 39 岁的黎曼病逝，留下一些未发表的手稿。结果，他的“尽职”

管家将这些手稿的大部分都付之一炬。剩下的一部分被黎曼的妻子抢救下来，并交托给黎曼生前的朋友——数学家戴德金。然而，数年之后，黎曼的妻子后悔了，她担心这些手稿中有个人隐私。于是在她的要求下，戴德金退回了部分手稿。最后剩下的部分，被存放在哥廷根的图书馆。20 世纪 30 年代初，西格尔有机会从图书馆借到了这些手稿。可以想象到，在这些手稿中，混乱、未成熟的思想与公式堆积在一起，真如天书一般。1932 年，在经过不懈的努力后，西格尔终于从这一堆杂乱的纸片中理出了头绪。他意识到黎曼是在利用一个特殊公式来计算零点。而且黎曼的这个秘密公式，可准确有效地计算黎曼假设中零点的位置，它远远胜过了当时数学界已知的任何计算方法。黎曼远远走到了整个数学界的前面，即便在他去世 65 年后，最著名的数学家仍然无法与之抗衡。而为了表彰西格尔在整理这一公式上的功绩，这一公式被恰当地命名为黎曼－西格尔公式。

西格尔的努力，还使黎曼的另一面得以展现。在此之前，黎曼被认为是具有非凡数学直觉的天才，而黎曼猜想的论文，也被数学家认为是黎曼个人非凡洞察力的体现。在哈代、朗道等数学家看来，这位依靠直觉和概念的数学家并没有为他的思想提供强有力的证据。然而，西格尔的发现，揭示出黎曼的论文是基于可靠的计算和理论思想的。

黎曼秘密公式的重新发现，大大推进了人们对黎曼零点的计算。几年后，哈代的一个学生用它成功计算出黎曼 ζ 函数的前 1041 个零点的位置，并证实它们都落在临界线上。这一借助一台打孔式计算机获得的结果首次突破了 11 年前 138 个零点的纪录，并标志着黎曼零点计算从纸笔到机器的转折。下一个将计算推进下去的是著名英国数学家图灵（1912—1954）。由于受到哈代的悲观态度影响，图灵也倾向于认为黎曼猜想或许是错的。为了说明这一点，只需要找到一个反例，即找到一个不在临界线上的零点就可以了。被誉为“计算机与人工智能之父”的他清楚机器可以成为搜寻零点的最有力工具。经过努力，在 20 世纪 50 年代，他完成了自己的机器，并于 1953 年计

算出了前 1104 个零点。当然，这些零点都在临界线上。因此，他构造反例的努力没有成功。

随着计算机时代的真正到来，黎曼公式的真正潜力得以充分显现。通过把这一公式变成计算机程序，人们得以探索黎曼世界中从未想象过的新世界，而黎曼零点的计算也进入了快车道。1956 年，人们计算了前 25 000 个零点；1958 年，纪录刷新为 35 337 个；1966 年，纪录为 250 000。在 1968 年的一次有名计算中，通过利用一些绝妙的理论技巧，纪录被推进到前 350 万个零点，并确认它们都落在临界线上。

在搜寻零点的路途中，数学家对黎曼猜想的态度再一次发生了改变。第二次世界大战末期，哈代去世时，人们曾普遍认为黎曼假设可能是错的。然而，到 20 世纪 70 年代时，大多数数学家已经相信黎曼猜想是正确的。这一心理转向，一方面与搜集到越来越多的计算数据有关，另一方面则来自于黎曼猜想之美。许多数学家从纯粹美学的观念出发，相信如果有零点落在临界线之外，就像是优美图画中的一个污点，或者"就像是你去听一场音乐会，所有的音乐家都以和谐的方式演奏乐曲，突然大号发出一个很强的声音，掩盖了其他乐器的声响"。充满美的数学世界中不会出现这种丑陋的事情。然而，也有对黎曼猜想持怀疑态度的数学家对此不以为然。数学家唐・查吉尔就是其中著名的一位。

查吉尔无法接受仅仅从纯粹美学的观念出发，无视其他的支撑证据就相信黎曼假设正确的观点。对他而言，前 350 万个零点落在临界线上也不能提供足够的心理支持。因为他认为黎曼猜想的反例不可能出现在如此小的区域。当这位怀疑论者与另一位黎曼猜想的坚定信仰者邦别里相遇时，有趣的故事发生了。

邦别里（1940—　），是意大利著名数学家。



邦别里

在 1962 年攻读研究生时，邦别里就给出了素数定理新的初等证明，1974 年，34 岁的他荣获菲尔兹奖。他在数学上的成功，使他获得了普林斯顿高等研究院的职位。早熟的他在 15 岁首次读到黎曼假设的时候，就深深地被迷住了。并且这位“世界最博学、最特别的数学家”持有一种绝对的信心，相信黎曼假设是正确的。20 世纪 70 年代初，邦别里到查吉尔所在的研究所访问。谈到黎曼假设时，查吉尔对邦别里提出了挑战。“我在茶点时间告诉他目前还没有足够的证据来令我相信黎曼假设是正确的，因此我愿意和他对此事立下等额赌约。但这并不表明我认为黎曼假设错误，而是我愿意充当魔鬼的代言人。”双方约定以查吉尔认为具有判定性价值的 3 亿个零点为限。如果在前 3 亿个零点中，出现黎曼猜想的反例，就算查吉尔获胜。反之，如果前 3 亿个零点都落在临界线上或者黎曼假设被证明，则邦别里获胜。赌注是两瓶上好玻尔多葡萄酒。邦别里痛快地接受了挑战。

当时，计算机的能力还很弱，无法探索黎曼临界线的这一区域。但后来，计算机的速度突然得到了大幅提高。这时，有两个小组决定利用计算机全新的能力来计算更多零点。1982 年，2 亿个零点被宣布都落在临界线上。

按照常规，计算小组不会再费心多算 1 亿个零点，因为人们感兴趣的是数量级上

的推进，而把纪录仅仅提高 50% 并没有多大的意思。因此，当听到这一计算小组在 2 亿个零点处终止计算时查吉尔松了一口气，他感到自己可以再放心等待几年了。可惜，查吉尔的如意算盘没有打成。计算小组在查吉尔的一位好朋友那里听说了赌局的事。于是他们一鼓作气，又多花费了约 1000 小时的 CPU 时间继续把工作推进到了 3 亿，结果没有发现落在临界线外的零点。查吉尔输了。他带着两瓶酒找到了邦别里，邦别里和他分享了一瓶。查吉尔幽默地说，这也许是他喝过的最昂贵的酒。当时 1 小时 CPU 时间的代价是 700 美元，运行 1000 小时的 CPU 时间只是为了看到赌局的结果，这样一瓶酒的价值折合 35 万美元。

不过如此昂贵的酒也没有白喝，这件事使查吉尔成了黎曼假设的忠实信徒，因为在他看来，已经有充足的证据支持黎曼假设的正确性了。

随着计算机运算速度越来越快以及更好方法的使用，黎曼零点的纪录又获得了非常大的突破。2004 年 10 月 12 日，两位法国人宣布已对黎曼 ζ 函数前 10 万亿个零点做了验证，并证实它们都位于临界线上。

介绍至此，有一点有必要解释一下。当我们提到零点计算的时候，人们通常会认为是计算零点的数值。其实，除了最初那些小范围的计算外，上面介绍的大规模零点计算都不给出零点的具体数值，而是只限于验证零点是否在临界线上。相比而言，零点数值计算方面的进展要缓慢得多。比如，在 1982 年“计算”了前 3 亿个零点之后又过了几年，数学家才完成了前 2000 个零点真正的数值计算（精度达小数点后 100 位）。在零点真正的数值计算方面，除计算前面的黎曼零点外，数学家出于后面我们马上会提到的原因，还会做另一种选择，即取一个虚部很大的数，然后计算它附近黎曼零点的具体数值。在这方面，一项较近的结果是，两位法国人完成了对 10^{24} 附近 20 亿个零点的数值计算。

作为计算工具的计算机终于强大到可以在黎曼 ζ 函数世界中探索得足够远，并对那些可能产生反例的区域进行探索，但其作用的大小却与黎曼假设的对错直接相关。

如果黎曼假设是错的，计算机将有可能在探索黎曼假设的进程中起到非常重要的作用，人们可能真的通过它找到一个零点落在临界线之外，那么它将扮演“黎曼猜想终结者”的角色。但是如果黎曼假设是正确的，计算机的作用就小多了。因为它只能通过验证有限个零点落在临界线上，给出越来越多的证据，支持人们对黎曼假设的信心，但它却无法证明所有的无穷多个零点都落在临界线上，也就是说无论计算机能计算多少个零点，也无法以这种方式生成一个证明。

计算不等于证明。数学需要的是证明。上面我们已经介绍了数学家们在证明黎曼猜想中走过的一段艰辛路程。下面我们再来了解一下，最近几十年在这一著名猜想的证明方面所取得的新进展。

数学与物理的交汇

1997 年 4 月 7 日，一条特殊的新闻闪现在国际数学界的网络中。有人宣称证明了黎曼假设。这一信息最初是邦别里在 4 月 1 日的一封电子邮件中透露出来的。邮件中介绍说：“法国著名数学家阿兰·科纳在上周三于高等研究院所做的报告中提到了黎曼猜想的一些有趣进展。在他报告时，听众中一位年轻物理学家灵光一闪，突然明白了如何运用他的‘超对称费米－波色系统’来解决黎曼假设。经过 6 天不间断的工作，以及一种新的计算机语言 MISPAR 的帮助，这位年轻的物理学家最终解决了这个数学难题。”

在邮件转发的过程中，原本信件中存在的 4 月 1 日的字样消失了。于是，渴望

再次经历像证明费马大定理那样兴奋的数学界被邦别里的愚人节玩笑蒙骗了。也许有人会奇怪，为什么上当的数学家没有对邮件中提到物理学家证明黎曼假设这一点产生怀疑呢？实际上，异乎寻常地提到这一点正是邦别里恶作剧的高明之处。因为在 20 世纪 90 年代的数学界，作为数论核心问题的黎曼假设与物理学的紧密联系已众所周知。

事情要从 20 世纪 70 年代说起。

一位名叫蒙哥马利的英国剑桥大学研究生思考了一个看似疯狂的问题：黎曼零点在临界线上是如何分布的？在黎曼零点是否都落在临界线上尚属未知的情况下，探讨其分布特点确实有些荒唐。然而，出乎意料的是，正是这一思考开辟了一条可能通往证明黎曼假设的大道。

在考察零点的分布时，蒙哥马利最初相信零点的分布是随机的，而一个零点之后很快会出现另外一个零点。后来他却意外地发现，零点的分布并非完全随机，而零点后面不会伴随着另外的零点，零点不产生汇聚现象，零点之间似乎存在一种互相排斥的趋势。1971 年，在对零点之间的距离做过分析后，他提出一种函数用以描述黎曼零点的“对相关”。这一函数说，每对零点之间的差遵循一种特殊的法则，该法则由公式 $1-\left(\frac{\sin \pi\mu}{\pi\mu}\right)^2$ 给出。蒙哥马利不清楚自己的发现是否是首创，另外他也非常急切希望自己的发现能有一个解释。为解开谜团，1972 年他访问了普林斯顿，去拜访数论大师塞尔伯格。当时的塞尔伯格在数论界的地位差不多相当于过去的高斯。他有一个习惯，也与高斯类似：这位坚持独自探索的人，把自己的许多研究结果尘封在抽屉里不发表。一位数学家曾回忆说：“一次我在普林斯顿作报告，塞尔伯格走过来说：‘我在 1954 年第一次证明你的定理时，就是这样想的。’他向我解释了他所做的工作，那比我的视角要好得多。所以很明显，他对此了然于心……非同寻常的是，35 年之后他竟然还能记得某一个细节。”

令蒙哥马利庆幸的一点是，塞尔伯格总算没有说出“啊，是啊，这个结果我早知道了”的话。但令他失望的是，塞尔伯格没能帮他解开其发现背后的含义。

见过塞尔伯格后，蒙哥马利去喝下午茶。这是普林斯顿一项重要的休闲方式，通过它，来自不同领域的学者可以交流彼此的想法，由此往往撞击出一些意想不到的智慧火花。与蒙哥马利交谈的是李特伍德的印度学生周拉，就是那位与塞尔伯格唯一合作写论文的人。两人聊的时候，周拉发现了旁边著名的博学物理学家戴森。经周拉引荐，蒙哥马利与戴森坐在一起。戴森热情地问他最近在干什么，蒙哥马利坦诚地谈起自己关于黎曼零点对之间间距可能行为的研究。当他说到关于间距分布的公式 $1-\left(\frac{\sin\pi\mu}{\pi\mu}\right)^2$ 时，戴森的眼睛发出了光芒。“这不正好是随机厄米矩阵特征值的对关联函数嘛！”

由戴森口中吐出的这个奇怪概念，虽然在当时的数学家眼中是完全陌生的，但对物理学家而言却已是老朋友了。它早已使用在量子物理学中，用来寻找刻画原子特征的方式，具体而言是预测当一个重原子被低能量中子轰击时原子核的能级变化。因此，戴森所发现的是两个表面上毫无关系的知识领域——量子力学和数论——之间的关系。

蒙哥马利与戴森在普林斯顿茶话会上的这次短暂会面已经成为数学上的一段佳话。但问题亦随之产生。首先，这种关系真的存在吗？其次，蒙哥马利发现的关于零点分布法则（也称蒙哥马利对关联假设）可信吗？再次，如果这种神奇的关系存在，那么它又预示着什么？最后，能否利用这种关系得到新的重要结果？特别是，能否通过这一渠道证明黎曼猜想？对这些问题的探讨，揭开了黎曼猜想研究的一个奇峰突起的精彩篇章。

先是在戴森的启发下，蒙哥马利查看了饵原子核的能级之间的间距，令他难以置信的是，能级与黎曼零点两者之间确实存在惊人的相似性。他惊讶地发现，自己预测的零点分布模式居然与量子物理学家在重原子核能级中发现的模式是一致的。在当时，

重原子核的能级模式已经获得实验的证实。然而，蒙哥马利对关联假设仍只是一个猜测。黎曼非平凡零点的分布真的与这一猜测相符吗？只有通过计算临界线上足够远处零点的具体数值，才能验证这些零点的行为是否真像他预测的那样。这正是在零点真正的数值计算方面，数学家们要计算虚部很大区域处黎曼零点具体数值的原因。

被黎曼零点与能级的美妙联系深深吸引的数学家安德鲁·奥德兹克利用他的超级计算机做了这件事情。1989 年，他得到了虚部在 10^{20} 附近 175 587 726 个零点的数值。结果发现，在这一范围处的零点间距分布与蒙哥马利的预测非常吻合。蒙哥马利的新观点终于有了令人信服的证据。然而，故事还远没有结束。

在使用一种新的统计方法检测黎曼零点和量子物理之间的联系是否真的存在时，奥德兹克注意到黎曼零点的数据中出现了一些令人困扰的矛盾。一位精通量子物理和混沌理论的专家贝里敏感地意识到可以摆脱这一矛盾的方法：用混沌量子系统作为解释素数行为的最佳物理模型。科学的三段伟大旋律——量子物理（极小世界的物理）、混沌（不可预测性的数学）和素数（算术的原子）——竟然如此奇妙地融合在一起了。而量子混沌与黎曼零点这两个相距如此遥远的领域竟然具有同样的模式！这一相似性的发现令数学家们极为惊讶与兴奋。但兴奋之余，人们不禁要问，在能级和零点之间找到的这一联系，是否可以导致一些更实际的进展，即我们能否从这种相似性中收获新的果实呢？

1996 年，素数定理证明 100 周年，许多世界顶尖数学家与物理学家在一家商业机构的赞助下共聚在西雅图交流对黎曼猜想的想法。在讨论到黎曼假设与量子混沌的关联时，普林斯顿大学的数学家萨那克向在场的量子物理学家提出挑战：用量子混沌和素数之间的相似性告诉数学家一些关于黎曼世界、数学界不知道的结果。同样，萨那克提供了一瓶好酒作为奖励。在座的数学家孔瑞恰好手中有一个非常特别的问题可作为检验的对象：黎曼 ζ 函数有一个特性，称为矩，利用矩可以生成一个数列。哈代和李特伍德曾证明数列中的第一个数应是 1，后来李特伍德的一个学生证明下一个数

是 2。在这次会议之前，孔瑞做了大量工作，猜测下一数应是 42，这是当时数学家对这一数列所掌握的全部信息。孔瑞为物理学家提出的挑战是，用量子物理中的类似术语来解释 42。结果两年后，由同一商业机构赞助的第二次会议上出现了极有戏剧性的一幕。

在这次会议上，孔瑞准备报告自己的一个推测。这个推测是他在第一次会议后，与另一位数学家合作经大量努力后得出的，这一推测是，序列中的第四个数是 24 024。然而到会的一位物理学家基廷报告的题目令孔瑞极为惊讶，因为这个题目表明，基廷已经得到了一个公式，由这个公式可以生成序列中的所有数。于是，孔瑞在基廷做报告前半信半疑地问他："你真的将它算出来了吗?"基廷说是的。一场真正的测试开始了。他们在一块黑板旁开始计算这个公式是否真的预测了这个序列中的第四个数。最终他们完成了计算。"当 24 024 出现在最后的结果中时，那种感觉真是不可思议。"孔瑞回忆说。随后，在"经历了科学生涯中最兴奋的几秒钟"后，基廷冲进报告厅面对一群数学家讲述起数论。报告后，塞尔伯格被这个新想法吸引了。他发表意见说，这肯定是正确的。基廷告诉了数学家们以前从不知道的事情，同时赢得了萨那克的一瓶好酒。

自从伽利略和牛顿时代以来，物理和数学美妙结合的例子已屡见不鲜。第三章提到的黎曼几何与爱因斯坦广义相对论的结合即是其中经典的例子。然而，在黎曼零点与量子物理之间的完全出人意料的密切联系仍然令人惊叹不已，而两者相似性的发现带来了双重的收获。一方面，黎曼假设的证明可能在量子物理学中衍生出一些新结果。另一方面，人们可以用量子物理学的思想探讨黎曼世界的性质，特别是用这种思路来证明黎曼假设。事实上，这一关于黎曼假设证明的全新方法，现在已被许多人看作是最有希望成功的途径。正因为此，许多物理学家加入到寻找黎曼假设证明的队伍中来。这也是前面提到的邦别里的愚人节玩笑被那么多数学家信以为真的原因。不过如果数学上最重要的未解决问题真的被物理学家而不是数学家解决，那将是真正的惊世之举。

这是第一次用量子物理学的方法解决纯数学问题，是数学与量子物理学之间联系的非凡例证，也是奏响数学与物理交汇的又一伟大新篇章。

结　　语

本章中我们介绍的是涵盖面很广的素数问题。对此，我们只是择其要者（最重要的自然是黎曼猜想问题），介绍了数学家们在揭示素数秘密时走过的部分历程及在这一历程中收获的数学金蛋。

2000 多年前算术基本定理的证明，使人们认识到作为算术原子的素数的重要性。欧几里得的优美证明又确切地告诉人们，素数有无限多。其后，数学家们曾热衷于寻求真正有用、理想的能产生素数的公式。然而，沿着这条道路，人们的探索都失败了。但在素数问题的另一研究方向，即探讨素数的分布方面，数学家们却取得了部分成功。随着 18 世纪末高斯与勒让德的发现，素数研究的方向发生了重大改变。人们的兴趣转向对素数个数函数 $\pi(x)$ 的研究，并取得了许多新的进展。其中一项重大突破是证明了高斯、勒让德的猜想。于是高斯等人的猜想成了素数定理。更重要的发现则来自黎曼，其 8 页论文开辟了素数分布理论研究的崭新途径，彻底改变了理解素数神秘性的任务。“高斯最先听到素数音乐主旋律，而黎曼则发现了隐藏在素数杂音背后的旋律。”于是，我们看到，多个世纪以来，数学家凝视素数，期望从中发现某种神秘的规律。最终，黎曼站在一个全新的角度观察素数，通过引入黎曼 ζ 函数，他发现了完全出乎意料的规律，混乱分布的无序素数有望转变成严格有序的黎曼零点。为了刻画零点的有序性，

黎曼通过自己的计算与理论分析，提出了伟大猜想：黎曼猜想。

随着时间的推移，人们越来越深刻地体会到黎曼猜想的重要性。人们意识到，一系列重要的数论问题，特别是与素数有关问题的完满解决，都有待于位居素数研究核心的黎曼猜想的解决。人们确信，这一猜想的解决将使人们得到比现在所知多得多的有关素数的信息与结果。我们举其中一个重要的为许多数学家（包括黎曼）特别关注的问题为例。上面我们介绍过，高斯提出用 $\mathrm{Li}(x)$ 作为素数个数 $\pi(x)$ 的估计。一个由此引出的重要问题是，这一估计的误差是多少，是存在一个控制范围，还是存在比较严重的偏差？ 1850 年，切比雪夫证明了，对于足够大的 x ，两者之间的相对误差不超过 11%。后来数学家证明了，在估值方面最理想的结果是 $\pi(x)-\mathrm{Li}(x)=O(x^{\frac{1}{2}}\ln x)$ 。目前的研究结果距此还很远。但只要黎曼猜想成立，那么数学家所期待的这一最理想结果就马上可得证。这意味着，如果黎曼假设正确，那么不存在某个音符的声音比其他的都响，演奏素数音乐的乐队将会处于完美的平衡状态。这样，素数的秘密最终归结为黎曼猜想的证明，而这一猜想的证明将奏响美妙和谐的“素数的音乐”。

黎曼猜想的证明，使我们再次深深叹服于数学之美。早期数学家们曾为素数的无序、随机与缺乏规律而沮丧。然而黎曼猜想的成立表明，素数分布是相对有序的。于是，人们透过黎曼的眼睛看到貌似随机的素数分布背后存在着奇异的规律和秩序，这种规律和秩序就体现在黎曼零点的奇特分布中。无序的素数转变为特别有序的黎曼零点。而黎曼零点的有序性又恰好可以解释素数分布的表面无序。素数的秘密被揭开后，里面竟隐藏着如此美妙的风景。

此外，数学家们发现，黎曼函数的主题一次又一次地在许多数学分支中出现。“当数学家在数学世界中寻找前进的方向时，似乎所有道路都不可避免地路过这个著名景点：黎曼假设。”在前进道路上碰到黎曼假设的数学家还发现，只有承认它的正确性才能继续前行。于是，在黎曼猜想尚未被证明的情况下，对这一结论正确性抱有强烈信心的数学家们已在它的基础上发展新的理论，即假设黎曼猜想为真，而去推出新的命

题。粗略的统计表明，在当今的数学文献中已经有超过 1000 条数学命题以黎曼猜想的成立作为前提。也正因为有如此多的结论依赖于黎曼提出的挑战，人们更愿称它为“假设”而不是“猜想”。因为“假设”一词所蕴含的强烈意义，就是数学家试图在此基础之上建立一套理论。相比而言，“猜想”一词仅仅表示数学家对这个世界行为方式的预言。这就是现在黎曼猜想更多地被称为黎曼假设的原因。于是，我们可以说黎曼假设现在已成为数学大厦中的必要部分。它的最终证明将为上千个命题的证明补上漏洞，使其真正升格为数学定理。当然，如果黎曼假设被推翻，那么建立在它上面的众多结论亦将灰飞烟灭。

由此，我们可以领会到数学家们所确信的事实：一旦黎曼假设得到证明，解析数论将从整体上发生翻天覆地的变化，数论的许多领域都会由此得到巨大进展。

根据前面介绍，我们还认识到，素数的重要性已远远超出了它们作为算术原子的基本功能。迄今为止，素数已经为我们打开了数学中看上去毫无联系的领域间的大门，数论、几何、代数、分析、逻辑、概率论、量子物理——所有这些都在追寻黎曼假设的过程中被联系起来了。我们再次看到，数学各领域之间存在着千丝万缕的联系，数学是一座统一和谐的大厦。而黎曼假设的证明，也将对与其相关的众多领域产生积极而重要的影响。

此外，黎曼假设的证明还与人们现代生活的重要组成部分，比如网络安全密切相关。原因在于，在几十年前，被认为与应用无关的最纯的数论已经堕入人间，特别是目前应用最广泛的 RSA 密码体系正是依靠素数创制的。其中密码的编制需要借助大素数的寻求，密码的破译则归结为如何快速有效地将正整数分解成素因子之积。如果黎曼假设正确，那么就会存在一个寻找素数的快速方法，利用它数学家可以非常迅速地定位一个百位或任意位数的素数，从而可以构造出更多 RSA 密码。反过来，数学家们也推测，如果找到了黎曼假设的证明，也许能为分解正整数提供一些新的途径。这正是商业界会关注黎曼假设研究的原因。

“黎曼假设是很多人都研究过的难题，人们知道它有很多结果，它位于一个著名的古老领域即素数研究的核心。所以，如果你是魔鬼，让某位数学家出卖灵魂以获得一个定理的证明，那么什么定理是大多数数学家愿意得到的？我想它是黎曼假设。”“如果问一位职业数学家，整个领域中唯一最重要的未解决问题是什么，你肯定会得到这样的回答：黎曼假设。”“黎曼假设并不是一般问题，它是纯粹数学中最重要的问题。它体现了我们不能理解的一种极其深刻而重要的东西。”“这是一流的难题，与费马大定理相比，我总觉得黎曼假设是真正重要的难题。”——这就是数学家们对黎曼假设的看法。

黎曼假设的重要性及其蕴藏的美使许多精力旺盛的出色数学家投入到证明这一命题的行列中。上面我们已介绍了在冲击黎曼猜想的过程中，人们做出的几次突破。然而这些突破离完全证明黎曼猜想仍是遥不可及。值得补充的是，黎曼在 8 页论文中曾以完全确定的口吻说自己可以证明绝大多数的零点位于临界线上！这是一个比目前已知的研究结果都强许多的结论！可惜论文中他完全没有提及这一证明的细节。除了 1859 年的论文外，黎曼还曾在一封信中提到过这一命题，追求完美的他解释说自己的证明还没有简化到适合发表的程度。直到他去世，人们没有见到他的这一证明。而希望从他的未发表论文中获得有用信息的企图也破灭了。

我们已提到过，他未发表论文的大部分极其遗憾地消失于“热心”管家的炉火之中。而剩下的部分后来又有一些被他的妻子索回，这其中有一本完成于 1860 年的小册子。当时在巴黎的黎曼，因为天气十分糟糕，所以他待在自己住处研究数学。而那一段时间，正是他发表其 8 页论文后的几个月。因此，人们猜测，这小册子中记录的内容非常可能与黎曼 ζ 函数有关，并且很可能有黎曼对其猜想的重要想法。然而，可惜的是，这本数学界极其渴望获得的小册子被黎曼妻子索回后，其去向就成了一个谜。由此，黎曼留下了另一个后人无法揭开的谜中之谜：他真的证明了“绝大多数的零点位于临界线上”吗？要知道，在经过近一个半世纪为数众多的杰出数学家的努力后，

人们已证明的结果仍离他的结果相距甚远。但这次与多数后代数学家投费马的反对票不同，许多数学家愿意相信，黎曼自称中的证明是存在的。因为人们从他留下的手稿中发现，其一些演算和证明哪怕是时隔几十年后被整理出来，仍然会大大超越当时数学界的水平。我们曾介绍过的黎曼秘密公式就是例证。于是，黎曼那些失踪的手稿与费马撩人的页边评注一样，成为数学中的传说。

1900 年，希尔伯特在其著名演讲中曾把黎曼假设作为第八问题“素数问题”的首要问题向数学界提出。到 20 世纪终结的时候，23 个希尔伯特问题都已经基本解决，而解决这些问题的杰出数学家就是所谓的“荣誉一族”。但是，“守护黎曼假设的城墙现在看上去也许有了些许的不同，但是仍和从前一样无法攻克”。

2000 年 5 月 24 日，马萨诸塞州剑桥的克莱基金会发起了一场颇具历史意义的竞赛。为迎接新千年的到来，由包括怀尔斯和阿兰・科纳等世界顶尖数学家组成的一个小组，在数不尽的数学问题中选出未解决的最困难、最重要的 7 个数学问题，作为千年难题向数学界提出挑战。任何能够解决 7 个数学难题之一的人，在专家认定其解答正确之后，都可以获得 100 万美元的奖金。黎曼假设再次成为 7 个千年难题中的一个。这是在希尔伯特于 1900 年列出的最重要数学挑战中，唯一一个在一个世纪之后又列入千年难题表的问题，这更增加了它的魅力。

毫无疑问，任何人只要证明或者反驳了黎曼假设，就会获得巨大的荣誉，并将赢得令人心醉的不朽名声，其名字亦将镌刻在人类文明的智慧墙上。

随着量子物理进入数论的神圣殿堂带来的兴奋，以及由此出现的大量关于黎曼假设的研究成果，数学家们对这一难题的热情开始高涨。1994 年，怀尔斯的成功也重燃了人们的希望，数学家发现那些历史悠久的难题是可能被证明的。这众多方面的结合，使黎曼假设进入了其研究历史上最活跃的时期。自 20 世纪末以来，越来越多的数学家打算攻克这一数学中意义最重大的未解决难题。

2018 年 9 月 20 日，年已 89 岁的英国著名数学家迈克尔·阿蒂亚爵士（1929—2019）宣称自己用一个“简单而全新”的方法证明了黎曼猜想。一石激起千层浪，这则新闻引起媒体的极大关注，并使黎曼猜想进入了大众视野。

9 月 24 日，阿蒂亚在 2018 海德堡获奖者论坛上展示了他对黎曼猜想的研究结果。他开始报告时，海德堡获奖者论坛的官网被挤爆，直播频道瞬间崩溃。截至 9 月 25 日下午 4 时 30 分，共有 43 685 名观众通过论坛官方渠道收看了阿蒂亚的演讲视频。人们对黎曼猜想的热情被点燃了。

究其因，一是黎曼猜想本身的难度与知名度，二是宣称者阿蒂亚是数学界绝对重量级人物。“他以在几何和拓扑领域的贡献而闻名。他获得了许多奖项，特别是菲尔兹奖、阿贝尔奖和英国皇家学会最高荣誉科普利奖。”

有些令人感到意外的是，阿蒂亚在 40 分钟的正式演讲中花了半个多小时介绍了黎曼的生平并回顾了黎曼猜想的历史。直到演讲的第 35 分钟，备受期待的黎曼猜想证明才出现在观众眼前。阿蒂亚用一张幻灯片展示了他如何用反证法“过程超简单”地证明这一“数学史上伟大的猜想”。一张 PPT，三分钟讲解，阿蒂亚“通过理解量子力学中的无量纲常数——精细结构常数，并将此过程中发展出来的数学方法用于解决黎曼猜想”。在之后的提问环节中，阿蒂亚声称“用我的方法，黎曼猜想已经被证明了”。

黎曼猜想真被证明了吗？与媒体的热烈反应形成对照的是，数学界自始至终保持着平静。

在过去的几年里，阿蒂亚已经做出过两次重大宣称了。2016 年，他给出“6 维球面上不存在复结构”的证明。2017 年，他自称将长达 255 页的法伊特—汤普森定理的证明简化成了 12 页。对于他的这两次声明，数学家表示怀疑和沉默，他的证明也未被任何期刊发表。

阿蒂亚近几年的这类经历、年近 90 岁的现状，以及黎曼猜想的难度，使数学界对阿蒂亚关于黎曼猜想的证明不可能抱有乐观态度。

在他的演讲与 5 页预印本证明发表后，少数愿意实名点评阿蒂亚证明的人都持否定态度。加州大学河滨分校的数学物理学家约翰·贝兹说，这个证明只是把一个引人注目的主张堆砌在另一个主张之上，缺乏任何关联论证或真实的证据。

著名数学家丘成桐就此事做出评论："大家都说这篇文章没有提供一般数学家要求的严格性的定理证明……有时候不完备的证明也会带有启发能力，但是我还没有看到这篇文章的启发能力。"

出于对阿蒂亚爵士的敬重，数学界不可能大张旗鼓地宣称他的错误，或明确指出他的"证明"完全不构成证明。于是大多数学家对此的态度是，保持缄默，拒绝就此事公开评论。

2019 年 1 月 11 日，英国皇家学会发布讣告称，数学家迈克尔·阿蒂亚爵士去世，享年 89 岁。去世前 3 个月，阿蒂亚"冒着一个年轻学者不敢冒的风险"，在年轻学者不会赌的地方赌了一把。但他的冒险没有成功。他关于黎曼猜想的证明，将随着时间流逝，在数学家互有默契般的缄默中不了了之。他把攻克黎曼猜想的接力棒留给了后来人。人类仍然需要等待，等待一个真正有效的"全新的想法"。

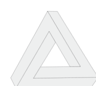
那么，迄今为止，数学家们离这座急欲攀登的数学之珠穆朗玛峰还有多远呢？谁将最终捧起这一纯粹数学的圣杯呢？

黎曼 ζ 函数的触角已伸向如此多的数学分支，甚至远至量子物理，因此人们不可能预测黎曼假设的证明会来自何处或何人。也许有一天某台计算机上会发现一个临界线之外的零点，从而推翻黎曼假设；也许真的有某位物理学家赢得这个数学头奖；也许黎曼假设的答案位于数学的核心深处，它需要由某位杰出的纯数学家来证明。事实上，在人们真正登上这座数学高峰之前，没有人知道哪条路能通向峰巅，也没有人知

道人们离峰巅究竟还有多远。也许证明这一猜想的所有拼图都已齐全，只等有人将其整合在一起。也许数学家解决这一难题所需要的数学工具或思想还未诞生。

因此，何时数学界可以为此举杯庆祝，对此我们没有答案。也许就在下一个月，也许要等 1 个世纪。假想 500 年后，复活的希尔伯特回到人间时，仍然没有人登上这座数学珠穆朗玛峰，捧起这一数学圣杯。

也许在对这个著名难题前景的猜测、期盼与无尽遐想中，关于数学问题的漫游戛然而止，这是最好不过的事情。



附录

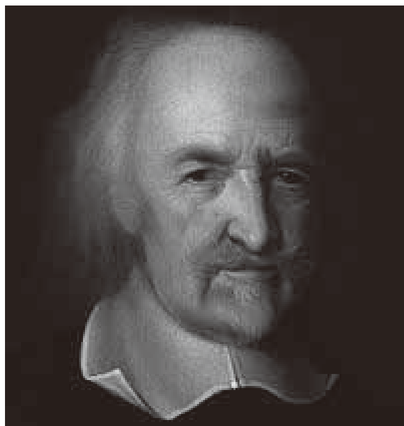
霍布斯与沃利斯——数学“民科”与数学家的一场较量

“民科”话题经常成为媒体关注与人们热议的焦点。本文提供了一则著名案例，详尽介绍了哲学巨匠霍布斯与著名数学家沃利斯的一场较量。文章展示了霍布斯在研究化圆为方等几何问题时数学“民科”的一面。在这场完全不对等的较量中，处于劣势的霍布斯至死不肯认输，不肯承认自己的低级错误。霍布斯的表现折射出“民科”的许多特征，这可以成为认识、理解“民科”行为与心态的一面镜子。

17 世纪中叶，一场旷日持久的争论拉开了序幕。这场争论的引人注目不仅在于它涉及一个著名的几何问题，更在于争论的双方都是当时（也是以后）知名的人物。其中一位是哲学史上大名鼎鼎的哲学巨匠霍布斯，另一位是 17 世纪数学史上极富创造力的数学家沃利斯。

哲学巨匠霍布斯：生平与著述

托马斯·霍布斯（1588—1679），是英国著名哲学家、思想家。



托马斯·霍布斯

霍布斯出生在英国威尔特郡一个乡村牧师家庭。4 岁时他就被送到当地的教会小学读书，后转到私立学校上学。霍布斯自幼聪颖好学，14 岁时已通晓希腊文和拉丁文。1603 年，15 岁的霍布斯以优异成绩进入牛津大学。但他的大学经历不是太愉快。当时的牛津大学给学生灌输的是经院哲学，科学在牛津根本没有地位，数学被当作魔术而加以禁止。他对学校中所教的正规课程不感兴趣，经常跑到附近的书店浏览地图和游记，把大部分时间花在其他事情上。但他最后还是以优异的成绩于 1608 年 2 月毕业，并取得了文学学士学位。

不久后，霍布斯经院长推荐去大贵族卡文迪许家任家庭教师。他的学生是比他年长 2 岁的男爵之子威廉，后来成为德文郡第二任伯爵。1614 年，作为家庭教师的霍布斯第一次陪同学生游历欧洲大陆。当时，英国贵族子弟为丰富知识，都要到法国和欧

洲大陆其他国家做周游旅行，人称“大周游”。对富家子弟来说，这种欧洲大陆游历几乎是一种惯例。

霍布斯的第一个学生只活了 38 岁。1628 年，亦生亦友的威廉去世后，霍布斯不得不改变原有的生活状态。1629 年，霍布斯成为格维斯・克利夫顿爵士儿子的家庭教师。不久就陪同新学生做第二次欧洲大陆游历。1630 年，霍布斯重新回到卡文迪许家担任威廉儿子德文希尔的家庭教师。1634 年，霍布斯陪同德文希尔又进行了为期两年的欧洲大陆游历。

出身于显贵家庭，特别是与卡文迪许家族建立的毕生友谊和联系使霍布斯找到了可靠的保护人和事业上的赞助者。通过卡文迪许家族，霍布斯也结识了当时英国一批有名望的社会名流与学者，如著名英国诗人和剧作家本・琼生。霍布斯结交的好友中还有著名的英国大哲学家弗朗西斯・培根。1620~1625 年，霍布斯曾给培根当了多年秘书。

陪同学生所做的三次游历更是对霍布斯思想的形成产生了极大的影响。

1610 年，霍布斯陪同学生第一次出游欧洲大陆，先后访问了法国、德国和意大利。这次大陆旅行，使霍布斯的眼界大为开阔。他意识到牛津所授各种课程已过时，世界需要重新认识。

第二次大陆旅行成了霍布斯走向哲学家生涯的转折点，从此，他把自己的兴趣从文学转向了科学和哲学。也是在这次游历中，霍布斯对几何学产生了兴趣。在一本他的早期传记中记载了一则有趣的轶事。一次霍布斯到一位法国绅士家作客，在主人的屋里，霍布斯发现桌子上有一本打开的欧几里得的《几何原本》。他恰好读到了“定理47”，他说：“这不可能！”于是他开始看对这个定理的证明。这个证明又引他去看前面的定理。最后，他信服了，并为几何学逻辑证明的严密和精确性所折服。这次富有传奇色彩的事情后，他爱上了几何学。

公理化的几何学方法第一次在霍布斯面前显示了自己的力量。在无可怀疑的基础上探讨事物之间联系的方法深深打动了霍布斯，并对其思想的形成产生了重要影响。

在他看来，自然科学必须采取几何证明的形式。必须有清晰的定义，公理必须被明白地规定出来，然后从中得出推论。只有这样，科学才是确定的、合乎逻辑的、必然的。几十年后，他熟练运用这种方法写出了自己的《论公民》。

第三次欧洲大陆游历是他思想发展最重要的一次旅行，给了他刚刚开始的哲学生涯深刻的影响。在这次旅行中，他结识了众多著名学者，特别是在法国认识了梅森，并与梅森组织的著名学术团体建立了联系。他还在 1636 年专程到意大利拜会了伽利略，与伽利略讨论了有关运动的各种问题。欧几里得的几何学给霍布斯以方法，伽利略的机械运动原理又给他以指导思想，一座哲学宏伟大厦的蓝图在他心中逐渐形成了。在他的计划中，他将完成的哲学体系包括三部分：第一部分是论物体，第二部分是论人，第三部分是论国家。

1637 年底，霍布斯怀着构筑体系的雄心返回英国。然而，不久后英国国内形势变得非常严峻。1640 年，在国王下令解散短期国会后，国内气氛达到了白热化。在这一时期，霍布斯用英文写了一本表明他政治理论的小册子《法律、自然和政治原理》。在这本书里他力图证明，国家权力不可分割地属于统治者，国王应该有绝对的权力。该书写成后，并没有出版，却以抄本的形式得到了广泛流传。

随着国会派和国王之间的斗争更加激烈，内战已迫在眉睫。霍布斯极为惊恐，他担心他的鼓吹君主专制的小册子会给自己带来生命危险。于是，在内战爆发前，霍布斯随同卡文迪什家族逃到法国避难。

流亡期间，霍布斯把大部分时间用于研究。1642 年，他把《法律、自然和政治原理》一书的后一部分做了扩充，最后完成并匿名发表了《论公民》一书。这本书出版后大受欢迎，就连笛卡儿也为之赞赏不已，霍布斯因而声名鹊起。1646 年初夏，霍布斯受到推荐，担任威尔士王子（即后来复辟后的查理二世）的数学老师直至 1648 年。

1651 年，霍布斯完成出版名著《利维坦》。虽然《利维坦》中几乎所有的中心思想都在霍布斯的早期著作或手稿中出现了，然而，相比他的前两本政治论著，《利维坦》的内容更为丰富，论证更为详尽，语气也更为激烈。这本书刚出版便产生了轰动。

很快，赞美与批评便蜂拥而至，使霍布斯的影响远远超出当时所有其他的思想家，霍布斯也由此赢得了渴望已久的声望并拥有了一大批崇拜者。不过，这本书的出版却使他与其他逃亡的保王派关系决裂，迫使他不得不向革命派的英国政府寻求保护。

1651 年，经过 11 年的流亡生活后，霍布斯回到克伦威尔执政的英国。在随后的岁月里，他在伦敦平静地从事他未完成的哲学体系的著述。1654 年，多年思索的成果《论物体》终于完稿，第二年在伦敦出版英文版。这本书着重论述了逻辑学、数学和物理学的基本原理。霍布斯把《论物体》看成是他的体系的第一部分，认为该书新阐述的基本原理是他的整个哲学体系的基础。1658 年，《论人》一书出版。至此，霍布斯计划完成的《哲学原理》三部曲（《论物体》《论人》《论公民》）全部问世，他终于完成了他构想了 30 多年的整个哲学体系。此后，霍布斯作为哲学家和政治思想家，没有写出更重要的作品，主要是捍卫自己的学说。

1658 年，克伦威尔病死。1660 年，查理二世即位，英国历史上开始了王政复辟时期。在其以前学生查理二世的保护下，霍布斯未受多大伤害，比较平静地度过了余下岁月。在晚年，霍布斯的兴趣又回到了青年时代的文学爱好方面。实际上，早在 1629 年，作为其研究文学和历史的第一个成果，他就曾翻译了古希腊历史学家修昔底德的《伯罗奔尼撒战争史》。1675 年，他将荷马的《奥德赛》译成英文，第二年又翻译了《伊利亚特》。1679 年 12 月 4 日，90 多岁高龄的霍布斯安静地离开了人世。“他的患病和去世没有什么特别与众不同的地方。”

然而，霍布斯的声望与影响并没有随着他的去世而烟消云散。他在世的时候，《法律、自然和政治原理》《哲学原理》（包括《论物体》《论公民》和《论人》）和《利维坦》这几本名著已为他赢得了国内与国际声誉，其思想影响了一大批著名学者，这些人中包括斯宾诺莎、莱布尼茨、狄德罗、卢梭、休谟和洛克。特别是体系最完备、内容最充实、论证最严密、学术价值最高、影响最大，被研究者称为“最伟大的，也许是唯一用英语写出的政治哲学巨著”《利维坦》一书为之后西方政治哲学的发展奠定了根基。这位政治理论的划时代人物因而也被称为第一位现代意义上的政治哲学家。第

二次世界大战后还出现了许多研究与复兴霍布斯思想的优秀著作。这反映了人类在日益复杂的生活面前，希望从他的思想中寻求启迪和答案的期盼。

由于思想上的分歧，霍布斯曾先后与许多同时代人展开过论战，如与著名法国哲学家笛卡儿、主教约翰·布兰豪等。这些论战的结果大都无损于霍布斯的声誉，而且有些争论或许还磨砺了他的思想。然而，很不幸，他一生中经历的最大争论。这一争论涉及霍布斯涉猎的另一方面：几何学。我们将看到，这位几何学研究者因沉迷于一个古老的几何问题而卷入一场不愉快的争论中，并因此丧失掉他在世时一度曾享有的数学家的声誉。下面我们先来看看卷入这场争论的另一方约翰·沃利斯。

数学强人沃利斯

沃利斯（1616—1703），是英国著名数学家、密码专家。



沃利斯

比霍布斯小 24 岁的沃利斯 1616 年 12 月 3 日生于英国肯特郡的阿什福德。在他年仅 6 岁时他受人尊敬的父亲因病去世了，但在母亲的照顾之下沃利斯接受了良好的教育。1625 年，沃利斯进入肯特郡坦特登村的一所文法学校，接受了完善的拉丁文训练。这个有天赋的孩子 14 岁时就能读和说拉丁文，还能读希腊文、希伯来文和法文。学校不教授数学，直到 1631 年的圣诞假期，15 岁的沃利斯才在兄长的指导下把算术当作一种“好玩的业余消遣”来学习。在首次和数学接触后，数学成为沃利斯的爱好。

1631~1632 年间，沃利斯就读于埃塞克斯郡菲斯台德村的马丁·霍尔拜克学校。除了进一步学习拉丁文和希腊文以外，还学了一点希伯来语以及逻辑学原理。1632 年圣诞节前后，作为自费生的沃利斯进入剑桥的伊曼纽尔学院攻读传统的大学课程，先后取得学士学位（1637）和硕士学位（1640）。在校期间，他学习过神学、物理学、解剖学、天文学及地理学等学科。

由于他所受的基本训练是当一名神职人员，因此在毕业后他担任了查理二世的宫廷礼拜堂牧师。但在业余时间他仍钻研数学，并做过一些数学研究，特别是在解代数方程方面获得了一些成果。能显示他当时水平的一个例子是，1647 年或 1648 年他重新独立发现了卡尔丹公式。在英国国会与国王查理一世间的冲突斗争中，他的数学才能也得以显现。他利用数学为国会破译了许多被截获的密码信件。

1649 年，颇具声望的牛津萨维尔几何教授职位出现空缺。6 月 14 日，沃利斯被意外地任命为牛津萨维尔几何教授，几乎可以肯定，这一任命是他为国会所做的有价值贡献的回报。自此以后，数学由他的业余爱好变成了职业需要，沃利斯也很快成为欧洲第一流的数学家。

直到 1703 年去世前的 54 年里，沃利斯一直保持这个职位。死后他被埋葬在牛津大学教堂，附近的一道墙上刻着下述文字：“这里安睡着约翰·沃利斯，神学博士，萨维尔几何教授，牛津大学档案保管人。他留下了不朽的著作……”

在著述颇丰的一生中，沃利斯完成并出版了多本数学著作，其中最著名的是《无

穷算术》(1655)。在这本名著中，沃利斯扩展了卡瓦列里的不可分原理，最早把指数概念从正整数推广到有理数，在对分数指数积分的研究中，求出了定积分（用现在符号表示）$\int_0^1 x^{p/q}\mathrm{d}x=\frac{1}{(p/q)+1}=\frac{q}{p+q}$；在求得定积分 $\int_0^1\sqrt{1-x^2}\mathrm{d}x=\frac{\pi}{4}$ 后，给出著名的无限乘积表达式 $\frac{4}{\pi}=\frac{3\cdot3\cdot5\cdot5\cdot7\cdot7\cdots}{2\cdot4\cdot4\cdot6\cdot6\cdot8\cdots}$。

这本代表作的发表，使沃利斯作为一个数学家享誉四方，其书中的发现，也使他成为微积分的先驱之一。几年后，这本书还深刻地影响了牛顿。1664~1665 年冬，正是在怀着极大兴趣阅读并深入研究《无穷算术》后，牛顿创立了微积分。

沃利斯另一本数学代表著作是《论圆锥曲线》(1655)。在书中，沃利斯第一次摆脱锥线是锥面截线的看法，最先把圆锥曲线当作二次曲线，并用笛卡儿引进的解析方法处理了这一经典主题。

此外，他还著有《一般数学或算术大全》(1657)，系统论述数学符号及历史概要。现在我们使用的无穷大符号“∞”与小于等于符号“≤”也是他的发明。1659 年，他写了《论摆线及蔓叶线》，将他所熟悉的解析法又往前推进了一步。1669~1671 年，他发表了长篇巨著《力学，或关于运动的几何学》。该书第一部分用严格的欧几里得几何方法讨论了各种不同形式的运动，在力学问题的数学化方面取得重大进展。1685 年，他发表了《论组合、交错与整除部分》一书，讨论了数论中的一些问题。沃利斯的最后一部数学著作是《历史的和实用的代数学》(1685)，这是第一本严格地叙述英国数学史的著作，全书分 100 章，对代数学的历史及其应用做了详尽阐述。在这部著作中，沃利斯还首先用几何方法解释了虚数。

作为 17 世纪最有才能和最有独创精神的数学家之一，也是牛顿在英国的直接前辈之一，沃利斯推动英国数学界的发展长达半个多世纪，在这段时间中，他为了促使数学在英国能享有与在欧洲大陆相同的显赫地位而做出了极大努力。正因为有了沃利斯的准备工作和牛顿的天才，数学研究的中心才从法国和荷兰转移到英国，并保持了一段时间。后来通过莱布尼茨、伯努利家族和欧拉的努力，这个中心才又移

回欧洲大陆。

除数学上的贡献外，兴趣广泛、智力超群、一生都精力旺盛的沃利斯还在其他多方面留下了自己的印迹。

沃利斯是伦敦皇家学会的主要发起人和坚定的会员。英国皇家学会成立于 1662 年，但早在 1645 年前后，沃利斯就参与了这一重要学术团体的创建。17 世纪 80 年代，沃利斯还当选为该学会的主席。这个起步时并不那么有名的以促进科学进步为己任的组织，后来成为声望极高的科学机构。

沃利斯还在密码、档案、语法、逻辑、音乐及神学领域有所建树。

作为一个密码专家，他早在青年时代，就为国会破译被截获的密码提供过有价值的服务。作为一位优秀的档案资料保管员，沃利斯所制定的编目分类法一直用到 20 世纪。他著有逻辑学著作 *Institutio logicae*（1687）。虽然不是职业音乐家，但他写出了一些关于音乐理论的论文，刊登在《哲学会会报》上。他写过不少宗教文章，他的布道及其他神学著作常因其通俗明朗的语言而倍受称道。

在语言学方面，他著有《英国语言文法》（1652）。在这方面的研究还为他教聋哑人说话的先驱性尝试奠定了一套有用的理论基础。他曾致力于训练耳聋者说话，据说是最先设计出一套教授聋哑人的方法的人。

沃利斯还以善于公开辩论而著称，他具有非常好争论的性格。与霍布斯一样，沃利斯一生曾多次卷入狂热的争辩。凭着自己卓越的数学才能，他屡屡得胜而归。例如，在 1656~1657 年间与法国著名数学家费马的论战后，沃利斯成了数学领域的权威。

他与争论最厉害的霍布斯的较量，会鹿死谁手呢？

霍布斯与沃利斯

霍布斯与沃利斯最初的争论源自一个几何学问题：化圆为方问题。这一问题是说，能否用直尺与圆规作出一个与给定圆面积相等的正方形。1654 年，霍布斯在自己出版的《论物体》一书第 20 章中对这一问题给出肯定的解答。1655 年，沃利斯发表文章，对霍布斯在《论物体》中“化圆为方”的理论表示质疑，争论由此开场了。

说起来，沃利斯为了发动这次对霍布斯的攻击已经等待了几年了。1651 年，霍布斯出版名著《利维坦》，为自己赢得了极大的声誉。但书中的一些观点也遭到来自不同阵营人士的反对。1654 年，牛津大学的天文学教授塞思・沃德（沃利斯的同事与朋友）发表《为学院辩护》为大学辩护，其反击的对象正是霍布斯。我们已提到，霍布斯的大学生活过得并不愉快，他因为厌恶大学里的神学教育，所以曾在《利维坦》中批评了大学，他指责牛津、剑桥大学是经院主义中心，是窒息科学发展的牢狱，并认为自己的书才应该“成为大学课本”。不满于霍布斯的傲慢与对大学的批评，沃德对霍布斯的观点做了驳斥，并为大学做了有力辩护。沃德声称，霍布斯根本不知道牛津大学课程的内容。他说，牛津的学术讨论是开放的，而非教条主义的；在数学方面，教的都是最新研究成果和方法。沃德批评霍布斯的观点就像“那 7 个沉睡不醒的人，很多年以后，他们醒来了，却依然按照当初入睡时的情形行事”。我们已经提到，1649 年，沃利斯成为牛津萨维尔几何教授。这之后，牛津大学在数学教学方面已经发生了很大改变。而霍布斯所谈论的是自己经历的大学，他不了解在其毕业几十年后一些大学的真实情况。因此，当霍布斯不顾变化后的事实，批评大学教育的时候，我们可以想见沃利斯的郁闷。

霍布斯《论物体》一书的出版，终于让沃利斯迎来了一扫这郁闷的机会。由于霍

布斯一直公开声称其哲学思想是以数学为基础的，因此，沃利斯希望通过指出霍布斯数学上犯的错误，以摧毁其哲学大厦。在后来（1659 年 1 月 1 日）写给荷兰物理学家和天文学家惠更斯的一封信中，沃利斯表明了这一策略。

“我们的利维坦先生正疯狂地攻击、破坏我们大学（不只是英国大学，而是所有大学）的名誉，特别是热衷于诋毁教士和牧师。似乎信教国家中没有任何真正的知识……似乎人们不懂哲学就无法理解宗教，而不懂数学又不能理解哲学。因此，正直的数学家看来有必要向他展示一下，他对自己口口声声所称的数学是多么无知。将他的推理倒过来，那么他向我们吐出污言秽语的所谓哲学，该是何等荒谬。”

于是，在《论物体》出版 3 个月后，沃利斯快速出击，向霍布斯发起攻击。在用拉丁文写成的小册子《关于霍布斯几何学问题的间接反驳》（间接反驳是苏格拉底采用的一种通过交谈而导出真理的方法）中，沃利斯批判了霍布斯的概念和方法，揭出霍布斯在数学上的许多漏洞。他以高超的论辩才能，时而冷嘲热讽，时而据理力争，把霍布斯描绘成一个鲁莽无知、狂妄自大的教会敌人。他对霍布斯的批判绝对称得上是无情揭露、残酷打击。

不过，“利维坦先生”可决非轻易服输之辈。

在 1655 年英文版《论物体》问世时，霍布斯虽然删除了一些被沃利斯揭发的严重错误，但他却增添了一个附录，对沃利斯的猛烈攻击进行了反击。这个辱骂性的附录题目是“数学六课：给一位几何学教授和一位天文学教授”。显然，这两位教授是指沃利斯和沃德。在献词中霍布斯写道：“从我《论物体》的第七章至第十三章，我已经修正和解释了科学［几何学］原理，即我做了沃利斯先生借以拿薪水的工作。”进而，他不仅回答了沃利斯的“间接反驳”，而且还提到了沃利斯的另外两本著作——“我在本书中仅用了两三页就完全将它们驳倒了。而且我坚信，自从世界产生以来，没有也将不会有比这更荒谬的几何学著作了”。

在第三课中，他将沃利斯的书称为“无知和莫名其妙的拼凑”。在第四课中，他将沃利斯的《无穷算术》一书称为“下流之作”（不过，正如我们前面已经提到过的，在

霍布斯眼中的这本“下流之作”促成了牛顿微积分的创建)。第五课特别引人注目，他批评了沃利斯书中“高度为无限小的平行四边形”。霍布斯问道:“这是几何学的语言吗?”其实，这恰恰反映了霍布斯的致命弱点：他过分痴迷于几何学而对迅速兴起的代数一窍不通。他本该客观地谈论沃利斯处理圆锥曲线的天才方法，但他却因为沃利斯的推导中“充满了如此之多的符号”而根本没有耐心去研究。

在结尾处霍布斯写道:“去你们的吧，粗鲁的教会主义、不近人情的牧师、道德沦丧者、狼狈为奸的朋党、以萨迦的一对坏蛋、最可恶的辩护者和所谓的学术捍卫者。”

沃利斯随即发表《为霍布斯先生改正错误——或曰校训(因为他的课没有讲对)》(1656)反驳霍布斯的观点，并趁着霍布斯发行《论物体》英文译本的期间继续批评他在数学上的错误。霍布斯则以数篇论文反击，如《荒谬几何学的标志、乡间俗语、苏格兰教会政治与几何学教授、神学博士亨利・沃利斯的愚昧》(1657)。但沃利斯轻易地以一篇回复(《驳霍布斯的观点》, 1657)击倒霍布斯的论点。最后霍布斯拒绝再回复沃利斯，暂时退出了争论，以完成他的三部曲的有关内容。沃利斯则利用这段时间整理出版了一本综合性文集《数学通论》，按现代术语讲，主要讨论了一些微积分的基本问题。两人的争论暂时停息。

经过一段时间的平静之后，霍布斯于1660年发表《对今日数学的检验和改进等》，宣布他又回到了争论战场。沃利斯选择了沉默。1661年，霍布斯又出版《关于物理学的对话》，目的是抨击波义耳和皇家学会的研究方法。波义耳作为沃利斯的朋友被当作了皇家学会的反面教材。在霍布斯看来，打击波义耳确可收一箭双雕之效。然而波义耳等人很轻易地便驳倒了他站不住脚的批评。

1662年，霍布斯又发表《物理学问题——兼论“倍圆”的两个命题》，开始在数学方面与包括沃利斯在内的几何学家展开正面交锋。这次涉及的几何问题除化圆为方外，又增添了“倍立方问题”。8月，霍布斯收到了惠更斯的一封信，信中解释了霍布斯在证明“化圆为方”和“倍立方”问题中出现的几处错误。另一位数学家看到霍

布斯论证“倍立方”问题的手稿复本后，也指出了霍布斯的错误，并对他又劝诫又侮辱：“这位杰出人物”在明显还没有“完全懂得数学家们的思想”时不应该对他们轻率地加以批评。如果霍布斯的“本来甚不错的名声”因为“老年时的这本可怜的著作”而大大受损了，他也会为之感到遗憾。霍布斯实际上被人当成了愚昧的老傻瓜。他对此很不高兴。霍布斯向所有人——那些“认为一方石头的十倍等于十方石头的人”除外——解释说，惠更斯等都是错的，而他是正确的。

沃利斯也写了一本《对霍布斯先生的〈对话〉的思考》（1662）批评霍布斯。霍布斯则写了一本《对霍布斯先生的忠诚、信仰、名声和作为的思考》作为回击。到1664年3月时，霍布斯厌倦了争吵：“我不想对已付印的证明再做什么改变、确认和争辩。它是正确的，如果那些怀有偏见的人没有仔细读它，那是他们的错，不是我的错。他们是一群自负的、喜欢在背后中伤别人的人。当他们在其他人的原理（这些原理要么是错误的，要么是不可理解的）上建构了错误的论证框架之后，他们的意识就为虚荣所缚，再也不愿承认任何新的真理。”

然后是暂时的休战。

经过一段时间的沉寂后，1666年霍布斯再次向沃利斯发起了猛烈的进攻，发动了第三阶段的争论，并且一直维持至他90岁为止。

1666年，霍布斯发表《几何原理及其证明》，打算以此击退自负的几何学教授的反扑。同时，他也坦诚地承认，自己正单枪匹马地与“所有的几何学家”作战。霍布斯还以其特有的幽默挖苦说：“或者我一个人疯了，或者只有我一个人没有疯，要不就是我们全疯了。除此以外，别无其他可能。”

1666年8月，沃利斯在皇家学会会刊《哲学学报》上发表了题为“写给一位朋友：沃利斯对霍布斯先生近作《几何原理及其证明》的评论”的文章，回答了霍布斯的挑战。

1669年，已是80多岁老人的霍布斯将解决化圆为方难题的种种方案以及对另外两个古希腊著名几何难题（用几何方法求球体体积、求倍立方体问题）的解法汇编出

版。该文集出版后，立即招来沃利斯的批评。两人于 1669 年、1671 年和 1672 年又爆发激烈争论。

晚年病痛缠身，但并不准备“安度晚年”的霍布斯在他生命的最后 10 年里，出版了三本大部头的几何学著作:《几何学的玫瑰园——此前未能证明而现已被证明的命题》(1671)、《因驳斥著名的牛津大学知名教授、神学博士约翰沃利斯而现出的数学之光》(1672)、《此前未能证明、现已得到简要解释和证明的几个几何学原理和问题》(1674)，另外还有去世前一年出版的物理学方面的著作《自然哲学的十篇对话》(1678)。每当自己的观点被沃利斯驳倒后，屡败屡战但绝不放弃的霍布斯就会出版回应批评的新版本，结果是沃利斯再次驳倒新的版本。

1678 年，90 岁高龄的霍布斯似乎仍意犹未尽，在他《自然哲学的十篇对话》中提出了论述物理问题的 10 个对话，为自己 30 多年来一直在提倡的几乎全部物理学理论进行了勇敢的辩护。同时，他还忘不了捎带批驳沃利斯的观点。

1679 年，霍布斯离开了人世。

在自己的一生中，霍布斯耗费了大量精力在几个几何作图难题的迷宫里东撞西碰。当我们指出他对几何学的看法后，就可以明白他所做的努力是何等徒劳。

在 1656 年的争论中，沃利斯批评霍布斯用希腊语的 stigma 表示“点”，他指出霍布斯本该用希腊语 stigme 表示数学上的没有大小、不占空间的“点”。但霍布斯却认为，这两个词可以互换，因为他不能想象存在没有大小、不占空间的点。同样，霍布斯还认为线段必须有宽度。

在霍布斯的观念中，几何学处理的是真实物质实体的问题。以“几何学方法本质上是要构造各种实体的图形”这一对几何学的误解为前提，霍布斯确信自己能解决“化圆为方”“倍立方”等几何难题。显然，无所畏惧的霍布斯没有意识到自己在这场争论中的真实地位。他虽然懂些几何，但他的数学修养远不能跟沃利斯相提并论。对懂数学的人而言，两人的争论谁是赢者显而易见。

然而，即便如此，在这场数学“民科”与数学家的看似完全不对等的较量中，处

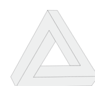
于劣势的霍布斯却一直没有认输，没有承认自己的低级错误。随着他的去世，两人这场持续了近四分之一个世纪之久的论战才落下了帷幕。数学强人沃利斯或许最终会意识到，训练耳聋者说话（他曾致力于此）也比让“利维坦先生”承认自己在数学方面的错误容易得多。

参考文献

[1] A. P. 马蒂尼奇. 霍布斯传. 陈玉明译. 上海：人民出版社，2007.

[2] 哈尔・赫尔曼. 真实地带——十大科学争论. 赵乐静译. 上海：科学技术出版社，2000.

[3] 韩雪涛. 数学悖论与三次数学危机. 北京：人民邮电出版社，2016.



参考文献

[1] 吴文俊. 世界著名数学家传记（上、下集）. 北京：北京科学出版社，2003.

[2] 梁宗巨. 世界数学通史（上册）. 沈阳：辽宁教育出版社，2005.

[3] 梁宗巨，王青建，孙宏安. 世界数学通史（下册・一）. 沈阳：辽宁教育出版社，2005.

[4] 梁宗巨，王青建，孙宏安. 世界数学通史（下册・二）. 沈阳：辽宁教育出版社 , 2005.

[5] 杜瑞芝. 数学史辞典. 济南：山东教育出版社，2000.

[6] 莫里斯・克莱因. 古今数学思想（一至四）. 张理京等译 . 上海：上海科学技术出版社，2005.

[7] 李文林. 数学史概论. 北京：高等教育出版社，2003.

[8] 李文林. 数学珍宝. 北京：科学出版社，2003.

[9] Victor J. Katz. 数学史通论. 李文林等译. 北京：高等教育出版社，2004.

[10] 柯朗等. 什么是数学. 左平等译. 上海：复旦大学出版社，2005.

[11] 张顺燕. 数学的美与理. 北京：北京大学出版社，2004.

[12] 张顺燕. 数学的源与流. 北京：高等教育出版社，2003.

[13] 基斯・德夫林. 数学：新的黄金时代. 李文林等译. 上海：上海教育出版社，1997.

[14] 单墫. 十个有趣的数学问题. 上海：上海教育出版社，1999.

[15] 张贤科. 古希腊名题与现代数学. 北京：科学出版社，2007.

[16] 徐诚浩. 古典数学难题与伽罗瓦理论. 上海：复旦大学出版社，1986.

[17] H. W. 伊弗斯. 数学圈（1、2、3）. 李泳译. 长沙：湖南科学技术出版社，2007.

[18] 马里奥・利维奥. 无法解出的方程——天才与对称. 王志标译. 长沙：湖南科学技术出版社，2008.

[19] 约翰・塔巴克. 代数学. 邓明立等译. 北京：商务印书馆，2007.

[20] 约翰・塔巴克. 几何学. 张红梅等译. 北京：商务印书馆，2008.

[21] 张景中，任宏硕. 漫话数学. 北京：中国少年儿童出版社，2003.

[22] 邱贤中，沈宗华. 几何作图不能问题. 上海：上海教育出版社，1983.

[23] 沈康身. 数学的魅力（二）. 上海：上海辞书出版社，2006.

[24] 华罗庚. 华罗庚科普著作选集. 上海：上海教育出版社，1997.

[25] 靳平. 数学的 100 个基本问题. 太原：山西科学技术出版社，2005.

[26] 蒋声. 形形色色的曲线 . 上海：上海教育出版社，1999.

[27] 蒋声. 欧几里得第五公设. 沈阳：辽宁教育出版社，1988.

[28] 伊莱・马奥尔. 无穷之旅. 王前等译. 上海：上海教育出版社，2000.

[29] 柳柏濂. 染色——从游戏到数学. 上海：上海教育出版社，2000.

[30] 沙特朗，张萍. 图论导引. 范益政等译. 北京：人民邮电出版社，2007.

[31] 西蒙・辛格. 费马大定理. 薛密译. 上海：上海译文出版社，1998.

[32] 胡作玄. 350 年历程——从费尔马到维尔斯. 济南：山东教育出版社，1996.

[33] 基思・德夫林. 千年难题. 沈崇圣译. 上海：上海科技教育出版社，2006.

[34] 卡尔・萨巴. 黎曼博士的零点. 汪晓勤等译. 上海：上海教育出版社，2006.

[35] 马科斯・杜・索托伊. 素数的音乐. 孙维昆译. 长沙：湖南科学技术出版社，2007.

[36] 韩雪涛. 数学悖论与三次数学危机. 北京：人民邮电出版社，2016.

[37] 韩雪涛. 从惊讶到思考——数学悖论奇景. 长沙：湖南科学技术出版社，2007.



微信连接



回复“数学科普”查看相关书单



微博连接

关注@图灵教育 每日分享IT好书



QQ连接

图灵读者官方群I：218139230
图灵读者官方群II：164939616

图灵社区
iTuring.cn

在线出版，电子书，《码农》杂志，图灵访谈

好的厨师为了更好地理解烹饪、寻找灵感，除了埋头钻研厨艺，往往还会周游世界，了解世界各地的饮食文化、食材和料理方式。类似地，如果你喜欢数学或者正在攻读数学系，但又不太了解这门学科的历史与现状，那这本书会对你非常有帮助。你将会穿越到过去，和书中的众多数学家一道，在贯穿数学史的6个重要问题中不断探索，不断前行，收获许多看课本和刷练习题无法学到的知识。

—— 顾森（Matrix67）

《思考的乐趣》作者

问题始终是数学发展的生命源泉，如果一个学科没有问题了，其生命力也就枯竭了。韩雪涛的这本书生动地告诉我们，自古及今，数学问题怎样吸引着世界上优秀的人，用他们卓越的智慧、坚韧不拔的意志，披荆斩棘取得一个个惊人成就。这本书的故事性极强，很多数学家有血有肉的性格呼之欲出：伽罗瓦的桀骜不驯、欧拉如贝多芬一样在失明后的多产、怀尔斯的谦逊和不屈不挠、埃尔德什浪迹天涯般地与人合作……这些精彩的故事、优美的数学、深邃的思想，必将吸引广大读者，追随着优雅的魔笛声，踏入数学的神奇殿堂。

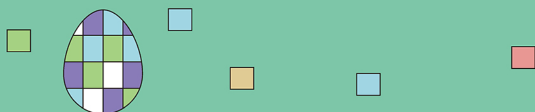
—— 刘新宇

亚马逊中国研发中心部门经理

“问题是数学的心脏。”一个好的数学问题不仅可以激发人类的智慧，而且还可以指引整个数学界前行的方向，探索更深邃的数学本质。在《下金蛋的数学问题》一书中，韩雪涛老师从数学轨迹中为我们采撷了6个经典数学问题，每一个都有自己的传奇：它的起源如何，在数学背景下为什么是重要的，开辟怎样的数学分支。我们能够透过了解这些问题迷人与伟大的意义，提高对数学的认识，锻炼头脑和思维的力量，激励自己像前人那样顽强学习，到达更广阔与自由的境界。

—— 李想

“遇见数学”公众号创办人



图灵社区：iTuring.cn
热线：(010)51095183转600
分类建议：科普读物 / 数学
人民邮电出版社网址：www.ptpress.com.cn



定价：79.00 元